Facultad de Ciencias

Departamento de Química Orgánica

Porous 2D Nanostructured Networks *via* Nucleobase Self-Assembly

Memoria presentada por

NEREA BILBAO BUSTINZA

Para optar al grado de

DOCTOR EN QUÍMICA ORGÁNICA

Madrid, 2016

La presente Tesis Doctoral ha sido realizada en el grupo de *Materiales y Sistemas Moleculares nanoestructurados* del Departamento de Química Orgánica de la Universidad Autónoma de Madrid bajo la dirección del Prof. David González Rodríguez.

Abbreviations and Acronyms

Common abbreviations and acronyms in organic chemistry have been used following the recommendations published by the American Chemical Society in their "Guidelines for Authors" (*J. Org. Chem.* **2015**).[1] Some other abbreviations have been used and are detailed below:

1D/2D/3D	one-/two-/three-dimensions
A	Acceptor (Hydrogen-bonding/Energy)
A	Adenine/Adenosine[2] – 2-aminoadenine/2-aminoadenosine[3]
AcOEt	ethyl acetate
AFM	Atomic Force Microscopy
AQ	anthraquinone
BODIPY	boron-dipyrromethene
C	cytosine/cytidine[2]
CB	central block
CCW	counter-clockwise
CD	Circular Dichroism
COF	Covalent Organic Framework
cor	coronene
CuPc	Cu (II)-phthalocyanine
CW	clockwise
D	Donor (Hydrogen-bonding/Energy)
DAN	2,7-diamido-1,8-naphthyridine
DAP	diacyldiaminopyridine
DBA	dehydrobenzo[12]annulene
DCC	Dynamic Covalent Chemistry
DCTB	*trans*-2-[3-(4-*tert*-butylphenyl)-2-methyl-2-propenylidene]malononitrile
DFT	Density Functional Theory
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DOSY	Diffusion-Ordered Spectroscopy
DSC	Differential Scanning Calorimetry
EM	Effective Molarity
eq.	equivalent
ESI	electrospray ionization
FAB	Fast Atom Bombardment
FRET	Förster Resonance Energy Transfer

[1] http://pubs.acs.org/paragonplus/submission/joceah/joceah_authguide.pdf
[2] The abbreviations G, C, A, U and T will be used both for nucleobases and their ribonucleosides derivatives.
[3] 2-aminoadenine will be abbreviated as A for the sake of simplicity.

FT-IR	Fourier Transform Infrared Spectroscopy
G	guanine/guanosine[2]
G_4	G quartet
GC-EI	Gas Chromatography-Electron Ionization
h	hours
HOF	hydrogen-bonded organic framework
HOMO	Highest Occupied Molecular Orbital
HOPG	Highly Oriented Pyrolytic Graphite
HRMS	High Resolution-Mass Spectrometry
$I/I_{set}/I_t$	tunneling current
iC	*iso*cytosine/*iso*cytidine[2]
iG	*iso*guanine/*iso*guanosine[2]
IUPAC	International Union of Pure and Applied Chemistry
K	association constant
LB	Langmuir-Blodgett
LDA	lithium diisopropylamide
LUMO	Lowest Unoccupied Molecular Orbital
M	molar (mol/L)
MALDI	Matrix-Assisted Laser Desorption/Ionization
MOCN	Metal-Organic Coordination Network
MOF	Metal-Organic Framework
MS	Mass Spectrometry
NB	nucleobase
NBI	naphthalene bisimide
NIS	*N*-iodosuccinimide
NMP	*N*-methyl-2-pyrrolidone
NMR	Nuclear Magnetic Resonance
NOESY	nuclear Overhauser effect spectroscopy
OA	1-octanoic acid
OP	oligo (*para*-phenylene)
OPE	oligo(phenylene-ethynylene)
OPV	oligo(*p*-phenylenevinylene)
OT	oligothiophene
OTf	triflate
P	porphyrin
pA	picoampere
PBI	perylene bisimide
Pc	phthalocyanine
POCM	Porous Organic Crystalline Material
Pp	phenyl porphyrin
ppm	parts-per-million

PTZ	phenothiazine
Q-TOF	quadrupole time-of-flight
RNA	ribonucleic acid
rt	room temperature
SAM	self-assembled monolayer
SOF	Supramolecular Organic Framework
STM	Scanning Tunneling Microscopy
T	thymine/thymidine[2]
TBAF	tetrabutylammonium fluoride
TCB	1,2,4-trichlorobenzene
TFA	trifluoroacetic acid
TGA	thermogravimetric analysis
THF	tetrahydrofuran
TMEDA	tetramethylethylenediamine
TMS	trimethylsilyl
TMSA	trimethylsilylacetylene
TTF	tetrathiafulvalene
U	uracil/uridine[2]
UG	ureidoguanosine
UHV	ultrahigh vacuum
UPy	ureido pyrimidine
$V/V_{bias}/V_t$	bias voltage
V	Volt
XRD	X-ray diffraction
ΔH	change in enthalpy
ΔS	change in enthropy

Contents.

Introduction.

1. Supramolecular Chemistry and Molecular Self-Assembly.

While traditional chemistry focuses on the covalent bond, supramolecular chemistry refers to chemistry beyond the molecule.[4] It deals with the complex entities formed by the association of two or more chemical species held together by non-covalent intermolecular forces (0.4–21 kJ/mol). Contrarily to a covalent bond, in these interactions electrons are not shared, but electromagnetic-like effects are rather concerned. These interactions include metal coordination, hydrophobic forces, electrostatic effects, π-π interactions and hydrogen (H)-bonding (Figure 1). In π-π interactions, for instance, an electron-rich π-system interacts electrostatically with another molecule or even another π system. When a metallic element acts as a Lewis acid and bonds non-covalently with a given ligand that contains one or more pairs of electrons that can be shared with the metal, the resulting force can even compete in strength with a covalent union.

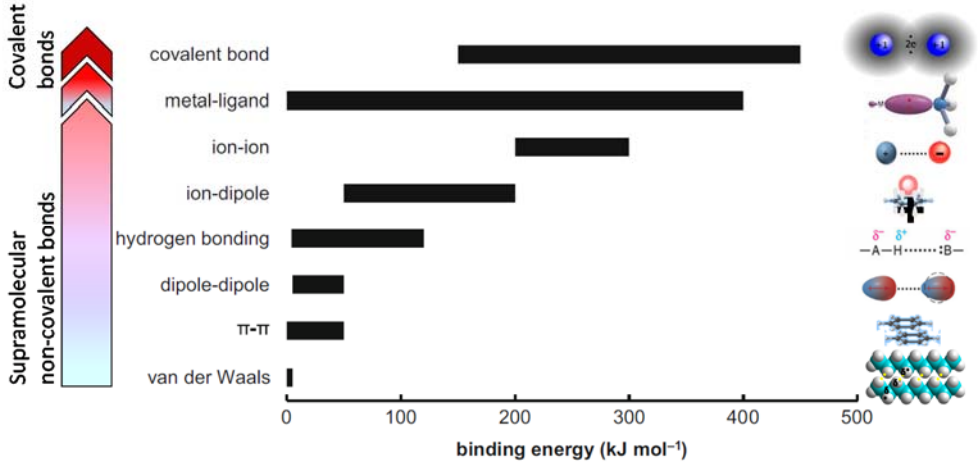


Figure 1. The most important non-covalent interactions along with their typical range of binding strength, in comparison with covalent bonds.[5]

In the last decades, supramolecular chemistry has widened the horizon of molecular self-assembly, folding, molecular recognition, host-guest chemistry, mechanically-interlocked molecular architectures and dynamic covalent chemistry (DCC). The latter relates to chemical reactions carried out reversibly under conditions of equilibrium control and lies between classic and non-covalent chemistry.[6] It offers a good illustration of how the thermodynamic control is central in supramolecular chemistry. Kinetically controlled reactions, resulting in the irreversible formation of strong covalent bonds (210–420 kJ/mol),

[4] J. M. Lehn, *Chem. Soc. Rev.* **2007**, *36*, 151–160.
[5] D. B. Varshey, J. R. G Sander, T. Friscic, L. R. MacGillivray, *Supramolecular chemistry: from molecules to nanomaterials* (Eds.: P.A. Gale, J. W. Steed), Wiley, New York, **2012**, *1*, pp 9–24.
[6] S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders, J. F. Stoddart, *Angew. Chem. Int. Ed.* **2002**, *41*, 898–952.

have been traditionally used in organic synthesis. Once a product is formed, it is usually not possible, under these same conditions, to transform it in another compound or to recover the starting material. On the contrary, in DCC the reactions occur under thermodynamic control and the products distribution depends only on their relative stability, and not on the relative magnitudes of the transition states. Therefore, the reversibility of the reactions in DCC allows some "proof-reading" process, where the most thermodynamically stable product is formed and the unstable ones are eradicated. By introducing certain features into the starting materials, a desired thermodynamic product can be attained and refined superstructures can be reached. For instance, one macrocyclic species can be selectively self-assembled under thermodynamic control using preorganized monomers. An interesting example was published by Rowan *et al.*[7] back in 1998, in which they formed cyclic trimers in > 90% yield from cinchona alkaloids, using transesterification as the thermodynamic cyclization reaction. The kinetic cyclization of a similar monomer gave, on the contrary, a mixture of cyclic products. Here, the difference in product distribution was attributed to predisposition of the monomer unit: the building block adopts a more stable conformation as a cyclic trimer than as a cyclic tetramer. Likewise, in supramolecular chemistry and molecular self-assembly, thermodynamics govern the outcome and allow some "error checking" of the process, so that the most stable species will be formed.

Although feeble, non-covalent interactions can become much stronger through cooperative and multivalent effects, as the forces can lean on each other to form stable structures. Cooperativity and multivalency are two distinct phenomena. The former describes how the binding of one ligand influences a receptor's affinity towards further binding interactions.[8] Cooperativity can be:[9] (1) positive (or synergistic) (Figure 2), when the subsequent binding is higher than the previous one; (2) negative (or interfering), when it is lower; and (3) non-cooperative (additive), when it is identical. A well-known example in biology is the allosteric oxygenation of hemoglobin (Figure 2a).[10] This tetrameric protein binds four individual oxygen molecules with increasing affinity, until the four binding sites are occupied, in a positively cooperative manner. However multivalency refers to the simultaneous binding of multiple ligands on one entity to multiple receptors on another. Nature frequently uses it to achieve tight binding in situations where univalent protein–ligand binding is weak.[11] In self-assembly, cooperativity takes a particular meaning, which includes the case of multiple intramolecular binding. Assessing cooperativity in multivalent systems requires the consideration of either (1) effective concentrations[12] of interacting groups within the multivalent receptor or (2) the additivity of free energies.[13] Positive

[7] S. J. Rowan, J. K. M. Sanders, *J. Org. Chem.* **1998**, *63*, 1536–1546.
[8] J. C. Badjic, A. Nelson, S. J. Cantrill, W. B. Turnbull, J. F. Stoddart, *Acc. Chem. Res.* **2005**, *38*, 723–732.
[9] K. A. Connors, *Binding Constants*, Wiley, New York, **1987**.
[10] W. A. Eaton, E. R. Henry, J. Hofrichter, A. Mozzarelli, *Nat. Struct. Bio.* **1999**, *6*, 351–358.
[11] a) R. T. Lee, Y. C. Lee, *Glycoconjugate J.* **2000**, *17*, 543–551; b) M. Mammen, S. K. Choi, G. M. Whitesides, *Angew. Chem. Int. Ed.* **1998**, *37*, 2755–2794; c) J. J. Lundquist, E. J. Toone, *Chem. Rev.* **2002**, *102*, 555–578.
[12] G. Ercolani, *J. Am. Chem. Soc.* **2003**, *125*, 16097–16103.
[13] P. I. Kitov, D. R. Bundle, *J. Am. Chem. Soc.* **2003**, *125*, 16271–16284.

cooperativity intervenes in making the supramolecular assembly progress toward a particular structure by encouraging the formation of a subsequent interaction, after the establishment of the previous one (Figure 2). The Gibbs free energy is more negative than the sum of individual forces, and in this way, the cooperativity of non-covalent interactions plays an important role in the engineering of sophisticated architectures.

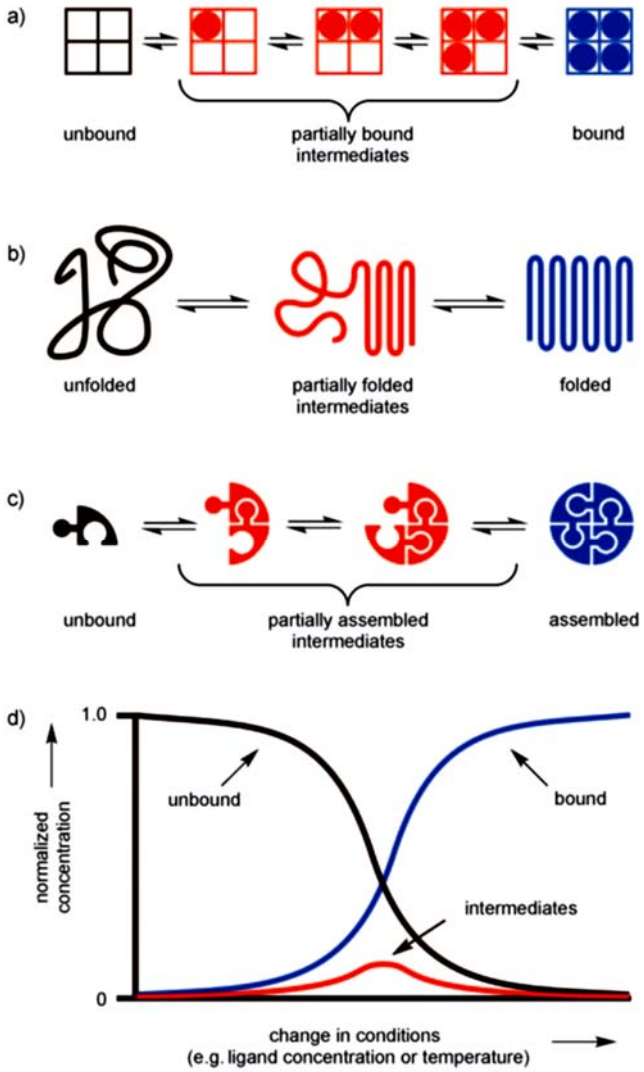


Figure 2. Representation of processes that display positive cooperativity: (**a**) hemoglobin binding oxygen, (**b**) protein folding, and (**c**) supramolecular self-assembly. (**d**) Speciation profiles. Positive cooperativity leads to a low peak concentration of intermediates and a sharp transition from unbound to bound species.[14]

[14] C. A. Hunter, H. L. Anderson, *Angew. Chem. Int. Ed.* **2009**, 48, 7488–7499.

Supramolecular chemistry brings physics and biology together with chemistry for the construction of highly complex species. Living systems assemble themselves and, with them, nature exhibits a large number of functional designs that suppose a boundless source of inspiration for chemists. Nature masters supramolecular chemistry and uses it to store, transmit and replicate information in a challenging environment and with a limited number of structural units. DNA and RNA are probably one of the best examples of this successful self-assembly. The four main natural nucleobases, meaning guanine (G), cytosine (C), adenine (A) and uracil (U)/thymine (T), are combined, in a cooperative and directional fashion, for the construction of these most interesting biopolymers. Also, DNA is an excellent illustration of a molecular recognition system: two complementary strands are linked together *via* non-covalent interactions into a 1D double helix. These forces are crucial to many biological processes and essential to comprehend how they work, meaning how the flow of information between natural structures functions. By extracting this expertise from biology, one can contribute to the solution of many problems in science and technology, in a non-biological context.[15]

The supramolecular chemist gets inspiration from nature and translates this expertise into the construction of molecular models that mimic natural systems through molecular self-assembly and the use of H-bonding, π-π stacking or van der Waals interactions. The design of "programmed" systems, controlled by molecular information, represents an interesting aim in materials engineering for the preparation of functional supramolecular materials, such as self-assembling nanostructures. When trying to achieve target nanostructures, one can carve a larger entity into different nanoscale devices by using externally controlled tools, such as micropatterning or nanopatterning. A top-down approach would be used in that case and corresponds to a miniaturization through the use of existing micro-/nanofabrication techniques. This method presents some limitations: on the one hand higher accuracies than 100 nm cannot be easily reached and, on the other hand, the level of organization is not optimal either. In contrast, in the bottom-up scenario, one can start from building blocks holding interesting physical and chemical properties and make them grow to form the desired conformation *via* self-organization or self-assembly.[16] Those starting single molecules carry the necessary information (size, shape and functionality) to provide the final structure equipped with the desired properties. Consequently, bottom-up approaches are competing with top-down strategies in providing new insights into nanotechnologies (Figure 3). The bottom-up approach deals with increased complexity at the molecular level and maintains an atom-by-atom control, where molecular self-assembly and molecular recognition hold an important role.[17]

[15] M. Boncheva, G. M. Whitesides, *Biomimetic Approaches to the Design of Functional, Self-Assembling Systems*, Marcel Dekker, Inc., **2004**, pp. 287–294.
[16] J. V. Barth, G. Costantini, K. Kern, *Nature* **2005**, *437*, 671–679.
[17] J. K. Gimzewski, C. Joachim, *Science* **1999**, *283*, 1683–1688.

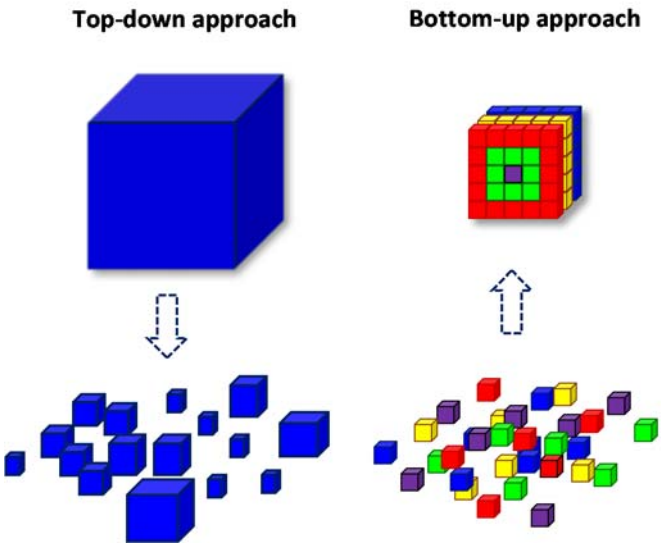


Figure 3. Illustration of the "top-down" and "bottom-up" approaches.

It has been briefly described in these paragraphs how biological systems use the cooperative and multivalent action of non-covalent interactions to construct and give function to the most sophisticated structures. Whereas molecular chemistry concerns entities constructed from covalently-bonded atoms, weaker forces are involved in supramolecular chemistry, where the spatial organization of these entities is governed by thermodynamics.[18] Great progress has been made in this field and among other phenomena, molecular recognition and host-guest chemistry hold an especially interesting spot. In this first part of the Introduction of this Doctoral Thesis, one section will be dedicated to H-bonding, making an insert on nucleobases. In a second section, the supramolecular construction of discrete macrocycles will be discussed and some important concepts will be introduced.

1.1. The Hydrogen-bond.

First introduced in 1935 by Bernal and Huggins,[19] H-bonds are electrostatic in nature and occur when a donor (*D*) group with a free acidic hydrogen atom interacts with an acceptor (*A*) group, carrying available non-bonding electron lone pairs. H-bonding is highly directional and selective and provides, therefore, exceptionally controlled geometric and spatial features. H-bonds show a maximum energy value for a 180^{o} angle and, accordingly, a linear binding is preferred. Two of the H-bonding energetic contributions are in the origin of this directionality: the so-called electrostatic or coulomb energy and the charge-transfer

[18] J.-M. Lehn, *Science* **2002**, *295*, 2400–2403.
[19] a) J. D. Bernal, H. D. Megaw, *Proc. R. Soc. Lond, A* **1935**, *151*, 384–420; b) M. L. Huggins, *J. Org. Chem.* **1936**, *1*, 407–456.

energy or covalent bonding. Both have an optimal angle of action. The strength of this interaction depends mainly on the solvent (it can compete for vacant H-bonding sites); the chemical nature of the H-bonding functions (an isolated H-bond is stronger when the basicity of the H-acceptor site and the acidity of the donor site are maximized); on the total number of H-bonds involved (the forces are additive) and their spatial disposition.[20] Also, the presence of intramolecular H-bonds, tautomerization phenomena, and the electronic effects of substituents can play a decisive role in the final strength of the overall interaction.[21]

Apart from the number of bonds engaged, the sequence of multiple H-bonds in a particular molecular fragment supposes an important parameter in its strength. Indeed, the geometrical disposition of the H-bond *D* and *A* functionalities is to be considered as a decisive factor. This phenomenon, with a strong influence in the stability of the H-bond-based complexes, has an electronic character. As Jorgensen and co-workers[22] showed, electrostatic attractive secondary interactions between positively and negatively polarized atoms in adjacent H-bonds increase H-bonding force. However, repulsive interactions between two positively or negatively polarized atoms lead to destabilization. As a result of these secondary interactions, the *DD–AA* motif is expected to be more favorable than the *DA–AD* motif.[23] For systems containing three hydrogen bonds: the *DDD–AAA* pattern is more stable than the *DDA–AAD*, which in turn is more stable than the *ADA-DAD* pattern. When exploring an example offered by nature, such as DNA base-pairing, one can see that the guanine-cytosine (G–C) pair, with 3 H-bonds, is far stronger than the adenine-thymine one, with only two H-bonds.[24] Even when considering 2-aminoadenine, the binding constant with uracil is significantly lower than between G and C, even if both pairs provide three H-bonds.[20e] In fact, the formation of the G–C dimer (*ADD–DAA* array) involves two attractive a two repulsive secondary interactions, whereas in the 2-aminoadenine–uracil dimer (*DAD–ADA* array), all the secondary interactions are repulsive, resulting in an association that is two orders of magnitude lower in $CHCl_3$. Generally speaking, complexes using the *DAD–ADA* motifs exhibit an association constant of around 10^2 M^{-1} in $CHCl_3$, whereas *ADD–DAA* complexes display binding constants of 10^4 M^{-1}.[25] *AAA–DDD* arrays can have association constants even higher than 10^5 M^{-1}. Sartorius and Schneider[23] proposed a simple empirical rule to predict the binding strength of a given complex. The free energy for dimerization could be calculated by adding a contribution of 7.87 kJ/mol for each H-

[20] a) R. P. Sijbesma, E. W. Meijer, *Chem. Commun.* **2003**, 5–16; b) L. Brunsveld, B. J. B. Folmer, E.W. Meijer, R. P. Sijbesma, *Chem. Rev.* **2001**, *101*, 4071–4097; c) J. L. Sessler, C. M. Lawrence, J. Jayawickramarajah, *Chem. Soc. Rev.* **2007**, *36*, 314–325; d) S. Sivakova, S. J. Rowan, *Chem. Soc. Rev.* **2005**, *34*, 9–21; e) S. K. Yang, S. C. Zimmerman, *Isr. J. Chem.* **2013**, *53*, 511–520.

[21] T. Marangoni, D. Bonifazi, *Organic Synthesis and Molecular Engineering*, (Ed.: M. B. Nielsen), John Wiley & Sons, Inc., Hoboken, New Jersey, **2014**, pp. 128–178.

[22] a) W. L. Jorgenson, J. Pranata, *J. Am. Chem. Soc.* **1990**, *112*, 2008-2010; b) J. Pranata, S. G. Wierschke, W. L. Jorgenson, *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.

[23] J. Sartorius, H.-J. Schneider, *Chem. Eur. J.* **1996**, *2*, 1446–1452.

[24] A. J. Wilson, *Soft Matter* **2007**, *3*, 409–425.

[25] T. J. Murray, S. C. Zimmerman, *J. Am. Chem. Soc.* **1992**, *114*, 4010–4011.

bond and ±2.9 kJ/mol for each attractive or repulsive secondary interaction. To sum up, the *DAD–ADA* complex has four repulsive secondary interactions, the *ADD–DAA* complex has two repulsive and two attractive interactions, and the *AAA–DDD* complex has four attracting interactions (Figure 4).

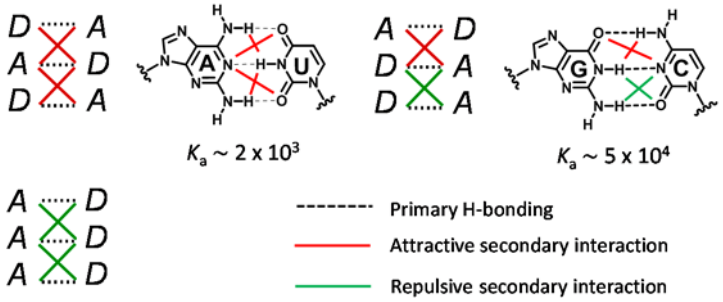


Figure 4. Secondary H-bonding interactions.

Apart from naturally occurring nucleobases, many research groups have develop new heterocycles that pair analogously (Figure 5). Additional H-bonds have also been incorporated to different building blocks in order to obtain interesting polymers and supramolecular structures with an increased stability and orthogonality.[20c,26] For instance, the group of Weck utilized cyanuric acid motifs,[27] which formed a three-point H-bonding pattern (*DAD–ADA*); or a six-point H-bonding array or Hamilton wedge (2 x *DAD–ADA*) associated with mono- or ditopic cross-linking agents based on 2,4-diaminotriazine, respectively. Binder and coworkers studied systems with T and diaminotriazine end groups (*ADA–DAD* array), a triple H-bonding motif with four repulsive secondary interactions.[28] Rotello and colleagues prepared functional systems with diacyldiaminopyridine and T[29] or U[30] derivatives, giving rise to a complementary three-point H-bonding (*ADA–DAD*). More interestingly, the scientific community has put much effort in developing stronger heterocomplementary arrays of quadruple H-bonds, serving to form supramolecular architectures (Figure 5).[31] Ureido pyrimidine (UPy) is one of the most important and extensively studied H-bonding motifs and a suitable one to engineer nanostructures.[32] UPy

[26] a) T. Park, E. M. Todd, S. Nakashima, S. C. Zimmerman, *J. Am. Chem. Soc.* **2005**, *127*, 18133–18142; b) E. M. Todd, J. R. Quinn, T. Park, S. C. Zimmerman, *Isr. J. Chem.* **2005**, *45*, 381–389.
[27] K.P. Nair, V. Breedveld, M. Weck, *Macromolecules* **2008**, *41*, 3429–3438.
[28] F. Herbst, K. Schröter, I. Gunkel, S. Gröger, T. Thurn-Albrecht, J. Balbach, W. H. Binder, *Macromolecules* **2010**, *43*, 10006–10016.
[29] F. Ilhan, T. H. Galow, M. Gray, G. Clavier, V. M. Rotello, *J. Am. Chem. Soc.* **2000**, *122*, 5895–5896.
[30] U. Drechsler, R. J. Thibault, V. M. Rotello, *Macromolecules* **2002**, *35*, 9621–9623.
[31] T. Rossow, S. Seiffert, *Supramolecular Polymer Networks and Gels*, (Ed.: S. Seiffert), **2015**, Springer International Publishing AG, Switzerland, pp.1–46.
[32] a) R. P. Sijbesma, F. H. Beijer, L. Brunsveld, B. J. B. Folmer, J. H. K. K. Hirschberg, R. F. M. Lange, J. K. L. Lowe, E. W. Meijer, *Science* **1997**, *278*, 1601–1604; b) J. H. K. K. Hirschberg, F. H. Beijer, H. A. van Aert, P. C. M. M. Magusin, R. P. Sijbesma, E. W. Meijer, *Macromolecules* **1999**, *32*, 2696–2705; c) N. E. Botterhuis, D. J. M. van Beek, G. M. L. van Gemert, A. W. Bosman, R. P. Sijbesma, *J. Polym. Sci., Part A: Polym. Chem.* **2008**, *46*, 3877–3885; d) B. J. B. Folmer, R. P. Sijbesma, R. M. Versteegen, J. A. J. van der Rijt, E. W. Meijer, *Adv. Mater.* **2000**, *12*, 874–878; e) H. Kautz, D. J. M. van Beek, R. P.

shows a strong dimerization in chloroform[33] as a result of the assembly of a self-complementary *DDAA* array of four H-bonds, pre-organized by an intramolecular H-bond. Besides UPy, Zimmerman and coworkers used 2,7-diamido-1,8-naphthyridine (DAN), which is a *ADDA* H-bonding array, and ureidoguanosine (UG), which has a *DAAD* pattern.[34] DAN and UG moieties allow the formation of robust structures, as a result of strong quadruple H-bonding between them, with four repulsive secondary interactions and two attractive ones. UPy can also interact with DAN by quadruple H-bonding in the form of its *ADDA* tautomer (Scheme 1).[33,35]

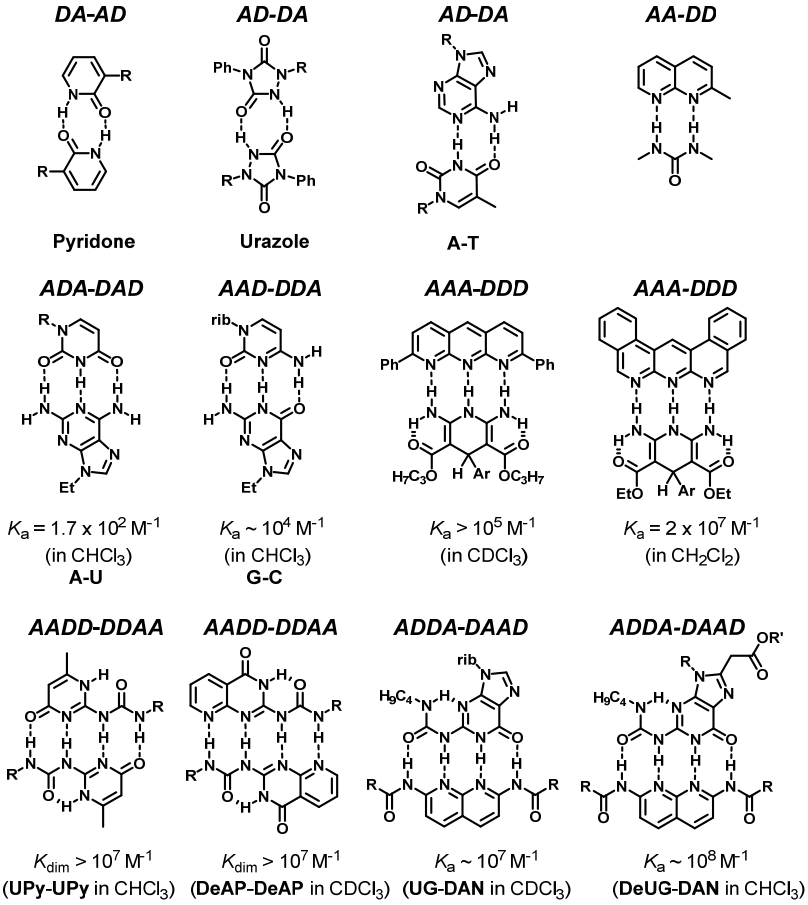


Figure 5. Some representative double, triple, and quadruple H-bonding motifs.

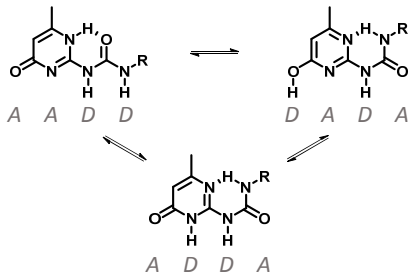
Sijbesma, E. W. Meijer, *Macromolecules* **2006**, *39*, 4265–4267; f) L. R. Rieth, R. F. Eaton, G. W. Coates, *Angew. Chem. Int. Ed.* **2001**, *40*, 2153–2156; g) K. E. Feldman, M. J. Kade, E. W. Meijer, C. J. Hawker, E. J. Kramer, *Macromolecules* **2009**, *42*, 9072–9081.

[33] F. H. Beijer, R. P. Sijbesma, H. Kooijman, A. L. Spek, E. W. Meijer, *J. Am. Chem. Soc.* **1998**, 120, 6761–6769.

[34] a) T. Park, S. C. Zimmerman, S. Nakashima, *J. Am. Chem. Soc.* **2005**, *127*, 6520–6521; b) T. Park, S. C. Zimmerman, *J. Am. Chem. Soc.* **2006**, *128*, 14236–14237; c) T. Park, S. C. Zimmerman, *J. Am. Chem. Soc.* **2006**, *128*, 11582–11590.

[35] X.-Z. Wang, X.-Q. Li, X.-B. Shao, X. Zhao, P. Deng, X.-K. Jiang, Z.-T. Li, Y.-Q. Chen, *Chem. Eur. J.* **2003**, *9*, 2904–2913.

As previously introduced, another parameter that can influence the H-bonding force is the tautomerization process that can occur when heteroaromatic systems are used as molecular recognition units. The possibility to have different tautomeric forms of the same molecule can induce an equilibrium between the *D* and *A* functionalities in the array, modifying their spatial orientation and, therefore, the strength and the specificity of the final interaction. UPy constitutes an example of this phenomenon, as it displays an equilibrium between three forms, where each tautomer displays a different H-bonding motif: *AADD*, *DADA* or *DAAD* (Scheme 1). The equilibrium between tautomeric forms weakens the association strength.



Scheme 1. UPy tautomeric forms.

We thought it would be pertinent to include a section about the H-bonding interactions in the specific case of DNA nucleobases, since they are crucial in our supramolecular design as self-assembling units. The interaction between the two anti-parallel DNA strands is primarily mediated by G–C and A–T complementary base-pairs, forming, on the one hand, the well-known double helix. RNA is, on the other hand, often a single-stranded molecule in many of its biological roles, such as coding, decoding, regulation and expression of genes, but its chemical structure is very similar to DNA. The unmethylated form of thymine, uracil, is the complementary nucleobase to adenine in the RNA biopolymer. These major binding motifs interact *via* two (A–T or A–U) or three complementary H-bonds (G–C), giving rise to the so-called Watson-Crick base-pairing. However, Watson-Crick pairing is not the exclusive mode of nucleobase association and 28 possible base-pairing motifs have been identified involving at least two H-bonds. These include reverse Watson-Crick, Hoogsteen and "wobble" (or mismatched) pairs (Figure 6).[36]

When H-bonding takes place on the Hoogsteen edge, the *N*-7 position of the purine base acts as H-bond acceptor and the C6 amino group as donor and the purine binds the Watson-Crick *N*-3 face of the pyrimidine base (Figure 6). Interestingly, a purine can undergo binding events both through their Watson-Crick and Hoogsteen edges at the same time. The Hoogsteen motif is, therefore, an additional H-bonding interaction available to purine

[36] M. Fathalla, C. M. Lawrence, N. Zhang, J. L. Sessler, J. Jayawickramarajah, *Chem. Soc. Rev.* **2009**, *38*, 1608–1620.

nucleic acids, already involved in Watson-Crick base-pairing. This multiple nucleobase interaction allows, for example, the formation of triplex DNA structures.

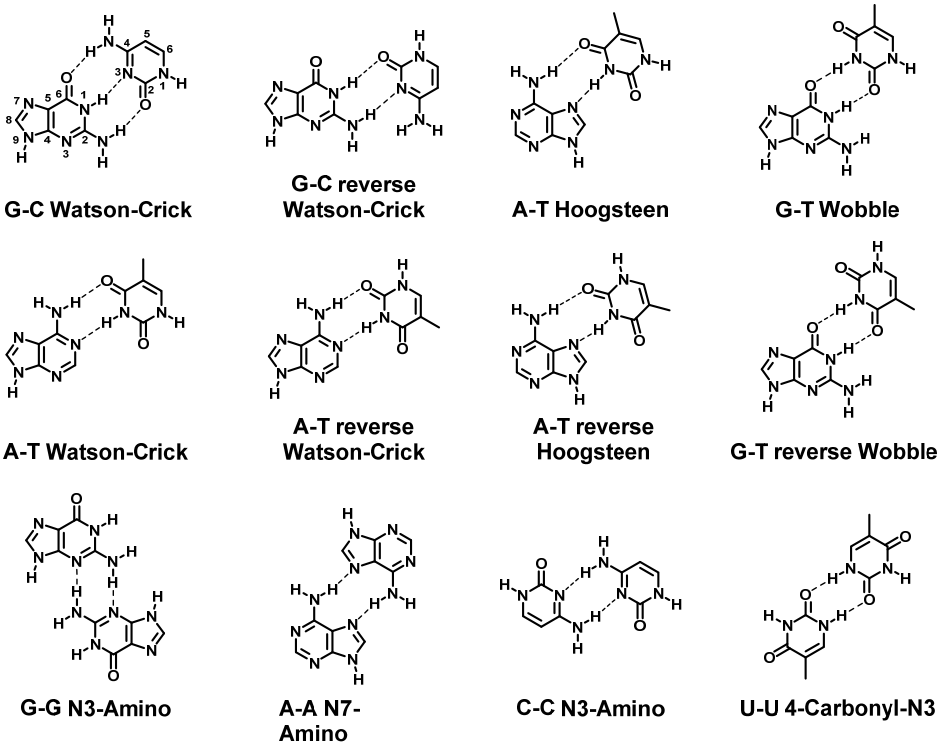


Figure 6. Watson-Crick, Hoogsteen and wobble base-pairs, as well as a series of homodimers assembled through two-point H-bonding interactions.

1.2. Hydrogen-bonded macrocycles.

When self-assembly is used to build a targeted well-defined (uniform) and discrete (monodisperse) structure, the term "supramolecular" or "non-covalent" synthesis is employed.[37] In order to achieve such a task, the design of the molecular entities must be exquisite. The self-assembling directors must have, encoded in their structure, the required chemical information needed to drive the association of the precise architecture; and thus, construct complex matter in the nano- and mesoscale. In this process, the term *fidelity* can be defined as the ability of a unit to produce a desired supramolecular complex, over all the possible ones.[26b] Also, in molecular recognition events, one can speak about *orthogonality*. It is imperative to understand individual non-covalent interactions, their synergy when several of them are present in the system, as well as how the overall system responds to environmental changes (solvent, temperature, concentration). Possible cooperative and

[37] a) L. J. Prins, D. N. Reinhoudt, P. Timmerman, *Angew. Chem. Int. Ed.* **2001**, *40*, 2382–2426; b) D. N. Reinhoudt, M. Crego-Calama, *Science* **2002**, *295*, 2403–2407.

multivalent phenomena between the individual blocks are of equal importance for a fine control on the shape of the structure and a well-defined function.[8,14,38]

A molecule with more than one binding site may in principle form open (linear) or closed (cyclic) structures (Figure 7). The formation of linear oligomers can be controlled by tuning nucleation-elongation processes[39] or the concentration. Also, end-capping units can be incorporated in order to quench the growth of the polymeric chain.[40] However, the supramolecular product is often a mixture of different systems with variable chain lengths.[20b,41] In order to reach well-defined and discrete supramolecular architectures, one approach can be to consider closed (multi)cyclic systems, where their size is determined by the geometric characteristics of both the monomer and the binding interaction. Templates leading to ring closure can also be utilized.[42]

Under thermodynamic control, cyclic species may be favored at the expense of oligomeric chains, since the intramolecular binding event may be preferred to the intermolecular process because of entropic and enthalpic reasons. This synergistic effect that causes such increased stability is defined as *chelate cooperativity*. Although other forms of cooperativity have been described (allosteric, interannular),[14,38] chelate cooperativity is the origin of many "all-or-nothing" processes, characteristic of discrete supramolecular systems, where either a given structure is formed or nothing else but the monomeric species can survive in a certain set of conditions.[43] A key parameter quantifying chelate cooperativity is the *effective molarity* (*EM*),[44] which affords an estimation of how favorable is the intramolecular binding interaction with respect to the intermolecular one. In other words, for thermodynamically controlled processes, the *EM* is defined by the ratio between the equilibrium constant leading to a cyclic system and the one leading to a linear system ($EM = K_{intra}/K_{inter}$).[45] Hence, the increase in stability when comparing a linear and a cyclic oligomer of a certain length (Figure 7) is given by the product $K_{inter}{\cdot}EM$, where the factor K_{inter} considers the additional association to form the cycle and *EM* takes into account that this last binding event is intramolecular.

---

[38] a) G. Ercolani, L. Schiaffino, *Angew. Chem. Int. Ed.* **2011**, *50*, 1762–1768; b) Focus Issue on Cooperativity. *Nat. Chem. Biol.* **2008**, *4*, 433–507.

[39] T. F. A. De Greef, M. M. J. Smulders, M. Wolffs, A. P. H. J. Schenning, R. P. Sijbesma, E. W. Meijer, *Chem. Rev.* **2009**, *109*, 5687–5754.

[40] a) S. P. Dudek, M. Pouderoijen, R. Abbel, A. P. H. J. Schenning, E. W. Meijer, *J. Am. Chem. Soc.* **2005**, *127*, 11763–11768; b) S. A. Schmid, R. Abbel, A. P. H. J. Schenning, E. W. Meijer, R. P. Sijbesma, L. M. Herz, *J. Am. Chem. Soc.* **2009**, *131*, 17696–17704.

[41] a) J. S. Moore, *Curr. Opin. Colloid Interface Sci.* **1999**, *4*, 108–116; b) F. W. Zeng, S. C. Zimmerman, S. V. Kolotuchin, D. E. C. Reichert, Y. G. Ma, *Tetrahedron* **2002**, *58*, 825–843; c) J. M. Lehn, *Polym. Int.* **2002**, *51*, 825–839; d) U. Michelsen, C. A. Hunter, *Angew. Chem. Int. Ed.* **2000**, *39*, 764–767.

[42] a) S. R. Seidel, P. J. Stang, *Acc. Chem. Res.* **2002**, *35*, 972–983; b) S. Sato, J. Lida, K. Suzuki, M. Kawano, T. Ozeki, M. Fujita, *Science*, **2006**, *313*, 1273–1276.

[43] J. K. Sprafke, B. Odell, T. D. W. Claridge, H. L. Anderson, *Angew. Chem. Int. Ed.* **2011**, *50*, 5572–5575.

[44] a) A. J. Kirby, *Adv. Phys. Org. Chem.* **1980**, *17*, 183–278; b) L. Mandolini, *Adv. Phys. Org. Chem.* **1986**, *22*, 1–111.

[45] a) X. Chi, A. J. Guerin, R. A. Haycock, C. A. Hunter, L. D. Sarson, *J. Chem. Soc., Chem. Commun.* **1995**, 2563–2565; b) G. Ercolani, *J. Phys. Chem. B,* **1998**, *102*, 5699–5703; c) G. Ercolani, *J. Phys. Chem. B,* **2003**, *107*, 5052–5057; d) G. Ercolani, *Struct. Bond.* **2006**, *121*, 167–215.

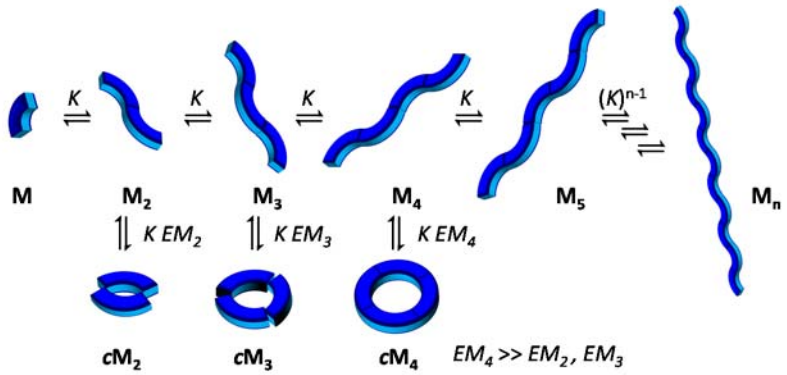


Figure 7. Possible supramolecular equilibria of a molecule with a given geometry and two binding sites associating with an intermolecular equilibrium constant *K*. Linear supramolecular oligomers (**M_n**) are in equilibrium with cyclic species (***c*M$_n$**). In this particular case, a cyclic tetramer is stabilized because the monomer and binding interaction geometric features afford a much higher *EM* value.

Enthalpic and entropic contributions in cycle formation regulate the magnitude of *EM*.[46] Its enthalpic component is mainly dependent on ring tension parameters and on attractive or repulsive alignment of molecular dipoles within the cycle. Therefore, the design of the monomeric unit has to be subtle and must take into account the preorganization of the structure toward a certain cycle size to achieve high *EM* values and hence quantitative yields of the target self-assembled structure. The entropic contribution is also determinant in the supramolecular macrocyclization. An intramolecular binding event may be preferred to an intermolecular one because some rotational and translational molecular degrees of freedom are maintained upon association, even if certain torsional motions are restricted in the closed structure and lead to a loss of entropy. When the monomer has several rotatable bonds and has a more flexible structure, this loss of entropy is magnified. Entropy also penalizes the formation of large cycles, where the *EM* tends to dissipate in closed assemblies formed from many molecules.

Among other non-covalent interactions that can drive the macrocyclization process,[45d,47] such as metal-ligand interactions, H-bonding arises arguably as the most interesting one since it allows a high degree of control over supramolecular architectures

[46] a) C. A. Hunter, M. C. Misuraca, S. M. Turega, *J. Am. Chem. Soc.* **2011**, *133*, 20416–20425; b) M. C. Misuraca, T. Grecu, Z. Freixa, V. Garavini, C. A. Hunter, P. van Leeuwen, M. D. Segarra-Maset, S. M. Turega, *J. Org. Chem.* **2011**, *76*, 2723–2732; c) H. J. Hogben, J. K. Sprafke, M. Hoffmann, M. Pawlicki, H. L. Anderson, *J. Am. Chem. Soc.* **2011**, *133*, 20962–20969; d) C. A. Hunter, M. C. Misuraca, S. M. Turega, *Chem. Sci.* **2012**, *3*, 589–601; e) C. A. Hunter, M. C. Misuraca, S. M. Turega, *Chem. Sci.* **2012**, *3*, 2462–2469; f) H. Adams, E. Chekmeneva, C. A. Hunter, M. C. Misuraca, C. Navarro, S. M. Turega, *J. Am. Chem. Soc.* **2013**, *135*, 1853–1863; g) H. Sun, C. A. Hunter, C. Navarro, S. Turega, *J. Am. Chem. Soc.* **2013**, *135*, 13129–13141.

[47] a) F. Würthner, C.-C. You, C. R. Saha-Möller, *Chem. Soc. Rev.* **2004**, *33*, 133–146; b) F. Biedermann, W. M. Nau, H. -J. Schneider, *Angew. Chem. Int. Ed.* **2014**, *53*, 11158–11171.

due to its high selectivity and directionality.[37a,48] Many research groups have studied various macrocyclization processes in solution by means of spectroscopic techniques (NMR, FT-IR, absorption, emission). As a matter of fact, the work by Reinhoudt and co-workers[49] on a melamine functionalized calix[4]arene represents one of the first examples of formation of discrete nano-objects through solely H-bond interactions. Upon mixing three equivalents of their calix[4]arene derivative (**1**-**3**) with six equivalents of the merocyanine dye (**BAR**), two coplanar chiral rosettes were formed (Figure 8). The assembly was held together by the highly cooperative action of 36 H-bonds.

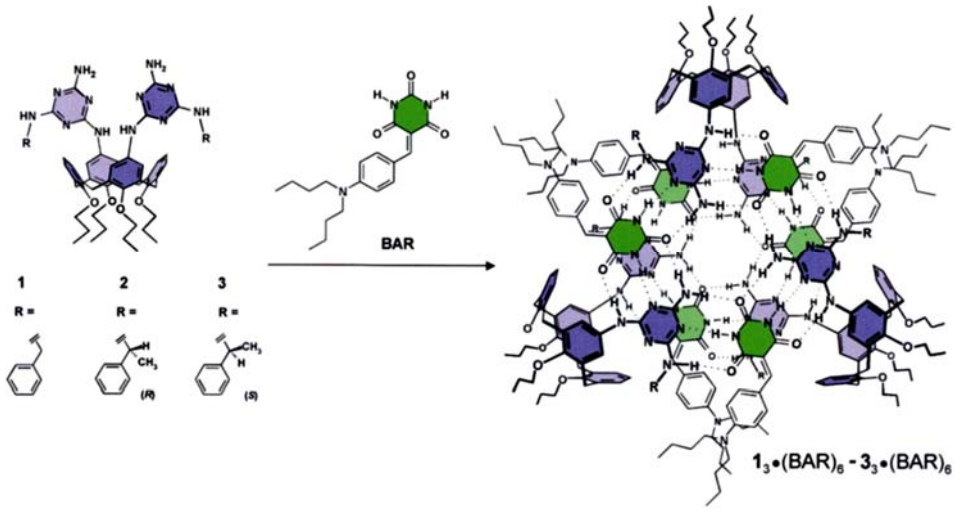


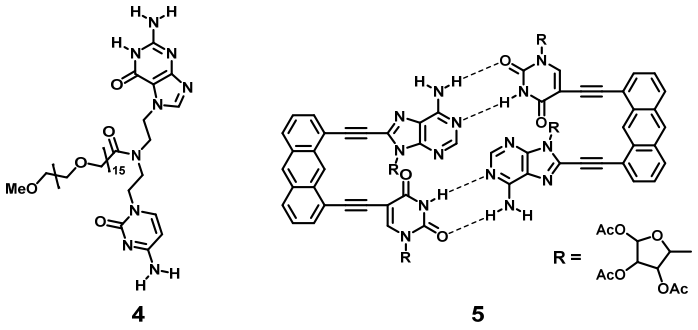
Figure 8. Schematic representation and molecular structure of the nine-component H-bonded assemblies and their individual components.

We will proceed below with a few selected examples from the literature, where the authors focused on the self-assembly of monocyclic H-bonded systems, from a single molecule with two nucleobase binding sites. These cases, from the formation of dimers to hexamers,[50] suppose an interesting precedent in "programming" discrete H-bonded macrocycles in solution that will pave the way to the self-assembly of similar DNA-based structures on the surface.

In the study of two-membered H-bonded macrocycles, Sessler *et al.* designed different ditopic molecules, based on complementary nucleobase pairs as recognition units, and evaluated their self-association depending on the rigidity of their system and by varying

[48] a) D. González-Rodríguez, A. P. H. J. Schenning, *Chem. Mater.* **2011**, *23*, 310–325; b) A. P. H. J. Schenning, D. González-Rodríguez, *Organic Nanomaterials: Synthesis, Characterization, and Device Applications* (Eds.: T. Torres, G. Bottari), John Wiley & Sons, Inc., Hoboken, New Jersey, **2015**, pp. 33–58; c) T. Steiner, *Angew. Chem. Int. Ed.* **2002**, *41*, 48–76; d) *Hydrogen Bonded Supramolecular Structures* (Eds.: Z.-T. Li, L.-Z. Wu), Springer-Verlag, Berlin Heidelberg, **2015**.
[49] L. J. Prins, C. Thalacker, F. Wurthner, P. Timmerman, D. N. Reinhoudt, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10042–10045.
[50] M. J. Mayoral, N. Bilbao, D. González-Rodríguez, *ChemistryOpen* **2015**. DOI: 10.1002/open.201500171.

the environment conditions (Figure 9).[51] In a first approach, ^{1}H NMR titration experiments of the self-association of **4**, compared to monomeric species containing G and C nucleobases (Figure 9) in DMSO-d_6 revealed a weak association constant for the dimer species due to the high conformational flexibility of the system, which implies a large entropic requirement to produce a ditopic complex and hence a low *EM* value.[51a] A second generation system was then prepared with a rigid skeleton.[51b] The strong self-association of **5** demonstrated that the rigid AU dinucleotide-based homodimers have an improved stability compared to the monomeric base pairs, whereas the use of flexible spacers hardly improved dimer stability. Solutions of a third generation rigid complex with bulkier substituents in $CDCl_3$ could be dissociated more easily than $\mathbf{5_2}$ by the addition of DMSO.[51c] When using a dibenzofuran moiety as rigid linker between the A and U nucleobases, the resulting dimer was even less stable due to a less optimized preorganization. A fourth generation rigid complex was prepared replacing the A and U bases with the stronger GC pair substituted with bulkier groups. Despite the fact that this dimer contained stronger H-bonds than $\mathbf{5_2}$, it was weaker than the AU dinucleotide.[51d] All these features indicated that either the lower structural preorganization or the presence of bulky protecting groups, which caused severe steric hindrance in the dimers, led to a significant reduction in the dimer stability. It is interesting to note that the kinetic behavior of the four systems was also strongly affected by their different geometric features and steric hindrances. While the more stable dimer $\mathbf{5_2}$ exhibited a slow equilibrium with the monomer in the NMR timescale, the rest of the assemblies were in the fast exchange regime at room temperature.



4 **5**

Figure 9. Dimeric structures based on nucleobase derivatives.

Years later, again the group of Sessler followed the previous strategy of designing artificial guanosine-cytidine dinucleosides to form cohesive dimers,[51d] but this time, to extend the supramolecular cyclization process to trimers (Figure 10).[52] They synthesized a derivative **6** able to form a trimeric structure through Watson-Crick base-pairing.

[51] a) J. L. Sessler, D. Magda, H. Furuta, *J. Org. Chem.* **1992**, *57*, 818–826; b) J. L. Sessler, R. Wang, *J. Am. Chem. Soc.* **1996**, *118*, 9808–9809; c) J. L. Sessler, R. Wang, *Angew. Chem. Int. Ed.* **1998**, *37*, 1726–1729; d) J. L. Sessler, R. Wang, *J. Org. Chem.* **1998**, *63*, 4079–4091.

[52] J. L. Sessler, J. Jayawickramarajah, M. Sathiosatham, C. L. Sherman, J. S. Brodbelt, *Org. Lett.* **2003**, *5*, 2627–2630.

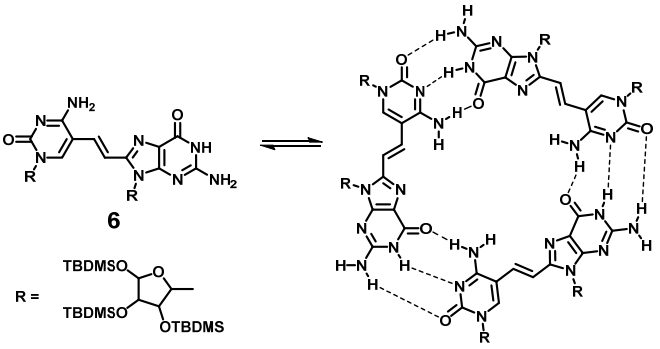


Figure 10. Self-assembly of **6** into a trimeric species.

Again, this potent GC H-bonding motif was used by Perrin *et al.*[53] to direct the association of compound **7** into tetrameric rosettes. The self-complementary heterocycle oriented the H-bonding faces of both guanine (*ADD*) and cytosine (*DAA*) at a 90° angle, which was dictated by the central pyrrole unit. Association then led to a tetrameric cycle containing 12 H-bonds (Figure 11). Variable-temperature ^{1}H NMR in DMSO-d_6/$CDCl_3$ verified H-bonding between the two faces of the monomer. The observation of new peaks at very low temperature was consistent with a G–C pairing scheme and suggested the formation of a tetrameric rosette in solution. DOSY experiments, as well as ESI-MS analysis, of **7** supported this tetrameric association, due to the fact that neither a peak for the trimer nor for any other higher-order aggregate were detected.

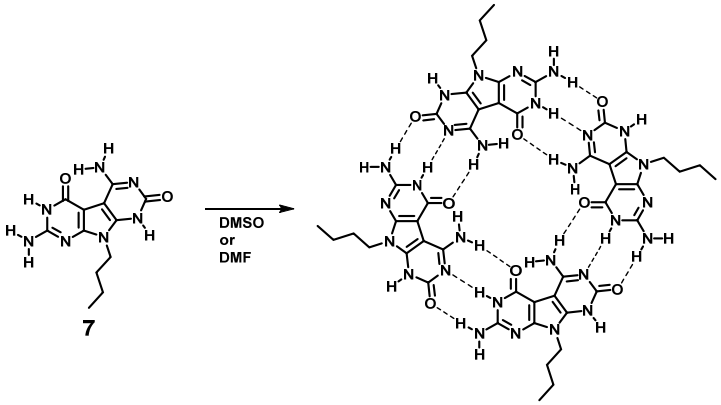


Figure 11. Self-assembly of **7** into a tetrameric rosette.

In 2011, Butkus, Wärnmark and colleagues reported the first selective assembly by tautoleptic aggregation of an enantiomerically pure cavity that is a supramolecular belt,

[53] A. Asadi, B. O. Patrick, D. M. Perrin, *J. Am. Chem. Soc.* **2008**, *130*, 12860–12861.

from one enantiomerically pure GC monomer containing one inherently non-self-complementary motif.[54]

The guanine molecule itself has a strong tendency to associate in tetrameric macrocycles leading to the well-known G-quartets (G_4),[55] where *DD–AA* double H-bonds are established through the Hoogsteen edges. This species that presents 4 carbonyl groups pointing toward the inner pore is further stabilized by cation complexation, typically with Na^+ or K^+ salts. Complexation leads, however, to multicyclic stacked assemblies, named G-quadruplexes, in which the cation is coordinated between pairs of quartets.[56] Since the stacked quadruplexes are the thermodynamic product, the quantitative formation of isolated G_4 has rarely been observed.[57] Cation-free or Na^+-bound quartets have been characterized in the crystal[57] or gas phases,[58] or using covalent[59] and non-covalent[60] templates. A few examples where the G_4 structure has been imaged onto surfaces are given in the final section of this Introduction.

The groups of Lehn[61] and Mascal[62] have pioneered the synthesis of Janus-type molecules like **8**, **9** and **10**, which contained H-bonding codes of both cytosine (*AAD*) and guanine (*DDA*) that mediated their self-organization into a hexameric supramolecular macrocylic structure (Figure 12). ^{1}H NMR spectroscopic and Vapour Pressure Osmometry measurements in chloroform indicated that the desired hexamer was formed in solution.[61] Moreover, X-ray crystallographic studies confirmed the presence of the cyclic hexamer $\mathbf{10_6}$ in the solid state (Figure 12b).[62b] In this structure, $\mathbf{10_6}$ hexamers overlapped each other to describe a highly porous solid with an interwoven network of channels (Figure 12c). This work represents one of the few examples where suitable monocrystals for X-ray analysis have been obtained and the resolved structure faithfully represented the assembly found in solution. The rigid nature and full preorganization of the molecule toward cyclization must inhibit the formation of linear oligomers and other H-bonded species and clearly helped in the formation of a well-ordered network at the solid state.

[54] E. Orentas, C.-J. Wallentin, K.-E. Bergquist, M. Lund, E. Butkus, K. Wärnmark, *Angew. Chem. Int. Ed*. **2011**, *50*, 2071–2074.
[55] a) J. T. Davis, *Angew. Chem. Int. Ed.* **2004**, *43*, 668–698; b) J. T. Davis, G. P. Spada, *Chem. Soc. Rev.* **2007**, *36*, 296–313.
[56] D. González-Rodríguez, J. L. J. van Dongen, M. Lutz, A. L. Spek, A. P. H. J. Schenning, E. W. Meijer, *Nature Chem.* **2009**, *1*, 151–155.
[57] a) J. L. Sessler, M. Sathiosatham, K. Doerr, V. Lynch, K. A. Abboud, *Angew. Chem. Int. Ed.* **2000**, *39*, 1300–1303; b) Y. Inui, M. Shiro, S. Fukuzumi, T. Kojima, *Org. Biomol. Chem.* **2013**, *11*, 758–764.
[58] C. Fraschetti, M. Montagna, L. Guarcini, L. Guidoni, A. Filippi, *Chem. Commun.* **2014**, *50*, 14767–14770.
[59] a) F. W. Kotch, V. Sidorov, Y.-F. Lam, K. J. Kayser, H. Li, M. S. Kaucher, J. T. Davis, *J. Am. Chem. Soc.* **2003**, *125*, 15140–15150; b) M. Nikan, J. C. Sherman, *Angew. Chem. Int. Ed.* **2008**, *47*, 4900–4902; c) M. Nikan, J. C. Sherman, *J. Org. Chem.* **2009**, *74*, 5211–5218; d) G. A. L. Bare, B. Liu, J. C. Sherman, *J. Am. Chem. Soc.* **2013**, *135*, 11985–11989; e) A. Laguerre, L. Stefan, M. Larrouy, D. Genest, J. Novotna, M. Pirrotta, D. Monchaud, *J. Am. Chem. Soc.* **2014**, *136*, 12406–12414.
[60] Y. Inui, S. Fukuzumi, T. Kojima, *Dalton Trans.* **2013**, *42*, 3779–3782.
[61] A. Marsh, M. Silvestri, J.-M. Lehn, *Chem. Commun.* **1996**, *13*, 1527–1528.
[62] a) M. Mascal, N. M. Hext, R. Warmuth, M. H. Moore, J. P. Turkenburg, *Angew. Chem. Int. Ed.* **1996**, *35*, 2204–2206; b) M. Mascal, N. M. Hext, R. Warmuth, J. R. Arnall-Culliford, M. H. Moore, J.P. Turkenburg, *J. Org. Chem.* **1999,** *64,* 8479–8484; c) M. Mascal, S. C. Farmer, J. R. Arnall-Culliford, *J. Org. Chem.* **2006**, *71*, 8146−8150.

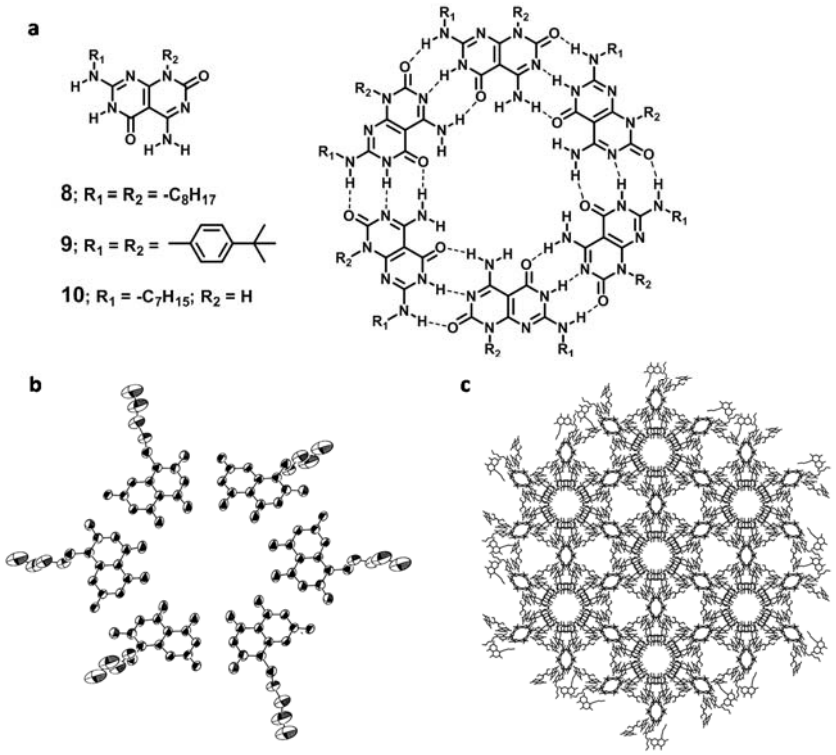


Figure 12. Supramolecular hexamer based on Janus-type molecules. (**a**) Chemical structure of the compounds **8**-**10**, along with the proposed self-assembly of the self-complementary heterocycles **8** and **9** and the self-assembled hexameric structure of **10**. (**b**) Oblique and (**c**) cell view of the hexagonal packing of **10** in the crystal.

The same motif was profited by Fenniri and colleagues to create a heterobicyclic GC monomer **11** that undergoes self-assembly in water,[63] as well as in polar solvents,[64] to form a six-membered macrocycle maintained by 18 H-bonds (Figure 13). The resulting hexameric rosette then self-organized hierarchically to produce tubular stacks (named rosette nanotubes, RNs) with tunable dimensions and properties. The peripheral diameter and its properties were dictated by the choice of the functional groups conjugated to this motif, whereas the inner space was directly related to the distance separating the H-bonding

[63] a) H. Fenniri, P. Mathivanan, K. L. Vidale, D. M. Sherman, K. Hallenga, K. V. Wood, J. G. Stowell. *J. Am. Chem. Soc.* **2001,** *123,* 3854–3855; b) H. Fenniri, B.-L. Deng, A. E. Ribbe, K. Hallenga, J. Jacob, P. Thiyagarajan. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 6487–6492; c) H. Fenniri, B.-L. Deng, A. E. Ribbe, *J. Am. Chem. Soc.* **2002**, *124*, 11064–11072; d) B.-L. Deng, R. L. Beingessner, R. S. Johnson, N. K. Girdhar, C. Danumah, T. Yamazaki, H. Fenniri, *Macromolecules* **2012**, *45*, 7157−7162.
[64] a) G. Tikhomirov, M. Oderinde, D. Makeiff, A. Mansouri, W. Lu, F. Heirtzler, D. Y. Kwok, H. Fenniri, *J. Org. Chem.* **2008**, *73*, 4248–4251; b) G. Tikhomirov, T. Yamazaki, A. Kovalenko, H. Fenniri, *Langmuir* **2008,** *24,* 4447–4450; c) A. Durmus, G. Gunbas, S. C. Farmer, M. M. Olmstead, M. Mascal, B. Legese, J.-Y. Cho, R. L. Beingessner, T. Yamazaki, H. Fenniri, *J. Org. Chem.* **2013**, *78*, 11421−11426.

arrays.[65] For instance, the authors designed a heterotricyclic self-complementary molecule **12**, which possessed the same H-bonding code, but was separated by an internally fused pyridine ring.[65] Compared to **13**, the assembled rosette nanotube **12** showed a higher diameter and a larger π-system that can allow electronic transport along the main axis. AFM images of randomly oriented RNs obtained from **12** featured an outer diameter of 4.3 nm, 0.9 nm larger than **13** (the inner diameter of **12** increased by 0.4 nm relative to **13**).[65b]

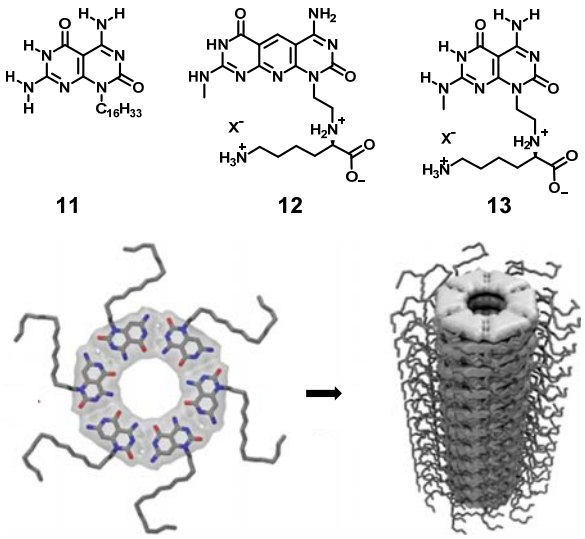


Figure 13. The self-assembly of heterobicyclic GC monomers. Chemical structure of the compounds **11**-**13**, along with the proposed self-assembly of monomer **11** into hexameric rosettes and RNs.

[65] a) G. Borzsonyi, R. S. Johnson, A. J. Myles, J.-Y. Cho, T. Yamazaki, R. L. Beingessner, A. Kovalenko, H.Fenniri, *Chem. Commun.* **2010**, *46*, 6527–6529; b) G. Borzsonyi, R. L. Beingessner, T. Yamazaki, J.-Y.Cho, A. J. Myles, M. Malac, R. Egerton, M. Kawasaki, K. Ishizuka, A. Kovalenko, H. Fenniri *J. Am. Chem. Soc.* **2010**, *132*, 15136–15139.

2. 2D Supramolecular Self-Assembly.

The engineering of materials with molecular-level precision requires adequate fabrication and characterization methods. Some sophisticated tools have been described to reach an exquisite control of the positioning of individual nano-objects, such as optical tweezers[66] or the tip of a scanning tunneling microscope.[67] Although attractive, these methods do not permit the preparation of a large number of nanosystems, as a profitable technological application requires. An approach for the fabrication of devices based on the self-ordering of objects at the nanoscale seems better-suited for applied nanotechnology. By ordering in an appropriate way molecules endowed with particular properties (catalytic, optical or electrical), the function of a given system can be optimized. The co-assembly of certain electron donors and acceptors for photovoltaic or optoelectronic applications is an example.[68] Additionally, the implementation of different materials into, for instance, electronic devices requires their controlled assembly onto a support, which can even sometimes play a bigger role and act as electrode. The deposition of organic molecules onto surfaces to make thin films of functional materials has, thus, become an essential part of numerous bottom-up fabrication processes. The nanoscale structure of such films determines the ultimate performance of the devices that they constitute. Therefore, it is crucial to deeply comprehend and have a real control on the growth and organization phenomena involved in their formation.

The most common approach for the preparation of 2D systems with a controlled geometry is to make the deposition of the molecules on the surface. The molecular modules should carry predefined recognition sites, able to induce the formation of an ordered network. The tools of supramolecular chemistry provide the guidelines to design molecular building blocks in order to achieve a balance between molecule-molecule and molecule-substrate interactions, as well as molecule- and substrate-solvent interactions when the latter is present, leading to the targeted functional patterns. The study and understanding of such systems could help in designing and constructing highly sophisticated artificial and functional 2D materials.[69] In this context, the continuous development of scanning probe microscopies has allowed the observation and characterization, at the molecular level, of the geometrical properties of 2D monolayers. Scanning Tunneling Microscopy (STM) arises as the most interesting technique to explore 2D supramolecular self-assembly, both under ultra-high vacuum (UHV) or at the solid-liquid interface. Additionally, the self-assembly of molecular building blocks on surfaces with well-defined cavities is of particular interest for

[66] S. M. Block, L. S. B. Goldstein, B. J. Schnapp, *Nature* **1990**, *348*, 348–352.
[67] a) D. M. Eigler, E. K. Schweizer, *Nature* **1990**, *344*, 524–526; b) R. Otero, F. Rosei, F. Besenbacher, *Annu. Rev. Phys. Chem.* **2006**, *57*, 497–525; c) L. Bartels, G. Meyer, K. -H. Rieder, *Phys. Rev. Lett.* **1997**, *79*, 697–700.
[68] a) P. Peumans, S. Uchida, S. R. Forrest, *Nature* **2003**, *425*, 158–162; b) L. Delgado, P.-A. Bouit, S. Filippone, M. A. Herranz, N. Martín, *Chem. Commun.* **2010**, *46*, 4853–4865; c) V. Percec, M. Glodde, T. K. Bera, Y. Miura, I. Shiyanovskaya, K. D. Singer, V. S. K. Balagurusamy, P. A. Heiney, I. Schnell, A. Rapp, H. –W. Spiess, S. D. Hudson, H. Duan, *Nature* **2002**, *417*, 384–387.
[69] K. Sakakibara, J. P. Hill, K. Ariga, *Small* **2011**, *7*, 1288–1308.

nanotechnological applications. Such "2D porous networks" offer the possibility to immobilize a variety of functional units as guest molecules in a repetitive and spatially ordered arrangement.[70] These nanopores can interact selectively with other molecular species in the environment, depending on their size and shape.[71] In this way, sophisticated applications can be developed, they may act as sieves, catalyst, chemical or biological sensors. In the latter, the selective adsorption of a given molecule on the self-assembled layer could, actually, lead to detectable electronic and/or structural modifications.[69,72]

In this coming section **2.1.**, different porous networks stabilized on the surface will be explained. Then, as in the previous part, the spotlight will be put on discrete self-assembled macrocycles, more closely related to the systems that are aimed in the framework of this Thesis (section **2.2.**). They will be discussed according to the different non-covalent interactions that were used to direct their self-assembly and form porous 2D architectures, such as van der Waals,[73] metallo-ligand interactions[74] and H-bonding.[75] The next section **2.3.** will be dedicated to the arrangements commonly adopted by nucleobases and related heterocycles, studied on solid substrates. In the final section of this Introduction (section **2.4.**), the topic chirality on the surface will be shortly introduced.

2.1. Porous networks.

Van der Waals forces arise from the attraction of the instantaneous dipoles generated by atomic electron clouds. Although weak, these interactions can stabilize different structures formed by molecules equipped with long aliphatic chains on highly oriented pyrolytic graphite (HOPG), where alkanes and aromatics have a strong adsorption. The interdigitation of alkyl chains from adjacent molecules is the main responsible of the stabilization of the 2D architecture. This event leaves the adsorption of the methylene hydrogens preferably into the centers of the graphite rings,[76] and stabilizes the assembly with an approximate energy of interaction of ~2.9 kJ/mol *per* methylene unit.[77] De Feyter, Tobe and co-workers[78] have made use of this approach to stabilize porous 2D-crystals from dehydrobenzo[12]annulene (DBA) (**14**) and its derivative **15** (Figure 14).

[70] a) T. Kudernac, S. Lei, J. A. A. W. Elemans, S. De Feyter, *Chem. Soc. Rev.* **2009**, *38*, 402–421; b) J. A. A. W. Elemans, S. De Feyter, *Soft Matter* **2009**, *5*, 721–735; c) X. Zhang, Q. Zeng, C. Wang, *RSC Adv.* **2013**, *3*, 11351–11366; d) V. A. Langlais, Y. Gauthier, H. Belkhir, O. Maresca, *Phys. Rev. B* **2005**, *72*, 085444.

[71] S. Stepanow, M. Lingenfelder, A. Dmitriev, H. Spillmann, E. Delvigne, N. Lin, X. Deng, C. Cai, J. V. Barth, K. Kern, *Nature Mater.* **2004**, *3*, 229–233.

[72] J. Mertens, C. Rogero, M. Calleja, D. Ramos, J. Martín-Gago, C. Briones, J. Tamayo, *Nat. Nanotechnol.* **2008**, *3*, 301–307.

[73] K. Tahara, S. Lei, J. Adisoejoso, S. De Feyter, Y. Tobe, *Chem. Commun.* **2010**, *46*, 8507–8525.

[74] A. Langner, S. L. Tait, N. Lin, C. Rajadurai, M. Ruben, K. Kern, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17927–17930.

[75] X. Zhang, Q. Zeng, C. Wang, *RSC Advances* **2013**, *3*, 11351–11366.

[76] J. P. Rabe, S. Buchholz, *Science* **1991**, *253*, 424–427.

[77] P. Samori, A. Fechtenkötter, F. Jäckel, T. Böhme, K. Müllen, J.P. Rabe, *J. Am. Chem. Soc.* **2001**, *123*, 11462–11464.

[78] K. Tahara, S. Furukawa, H. Uji-i, T. Uchino, T. Ichikawa, J. Zhang, W. Mamdouh, M. Sonoda, F. C. De Schryver, S. De Feyter, Y. Tobe, *J. Am. Chem. Soc.* **2006**, *128*, 16613–16625.

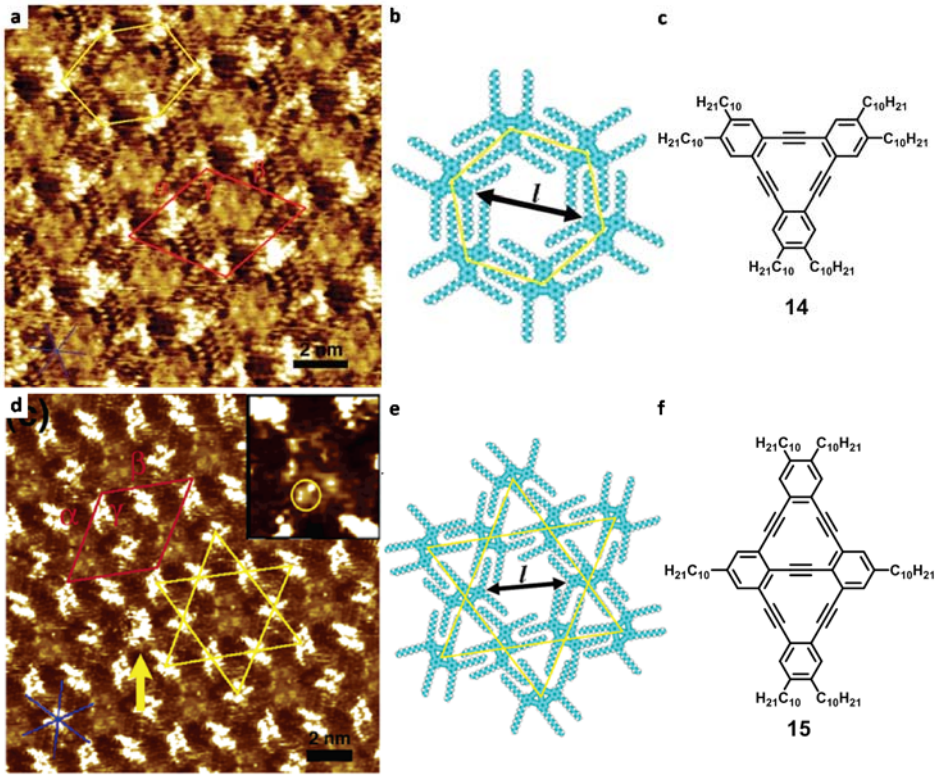


Figure 14. STM images of the different patterns adopted by the DBA-derivatives 14 and 15.(**a**) STM image of **14** physisorbed at the TCB/HOPG interface (I_{set} = 0.50 nA, V_{bias} = -0.99 V). (**b**) Molecular model of the honeycomb domain of **14** and (**c**) its chemical structure. (**d**) The Kagomé network of **15** (I_{set} = 0.50 nA, V_{bias} = -1.04 V). (**e**) Molecular model of the Kagomé network along with (**f**) the chemical structure of **15**.

The interaction between a metal and a ligand is among the strongest non-covalent forces. Therefore, the union of transition metal ions and organic moieties offer supplementary robustness to the supramolecular structures. Li *et al.*[79] formed on the Au (111) surface, under UHV conditions, a 2D porous assembly formed by the tetrapyridyl porphyrin **16** in the presence of copper (Figure 15a). Recently, Stöhr and co-workers[80] made also use of a pyridylporphyrin derivate (**17**) to self-assemble a cobalt-coordinated porous network on Ag (111) (Figure 15c). The group of Lin[81] used the same motif and were able to synthesize on the Au (111) substrate a large metal-bound porous self-assembly between a bisterpyridyl ligand (**18**) and the pyridyl-functionalized porphyrin (**16**), in the presence of Fe atoms *via* Ullmann coupling, reaching a pore size of around ~22 nm^2 (Figure 15e).

[79] Y. Li, J. Xiao, T. E. Shubina, M. Chen, Z. Shi, M. Schmid, H.-P. Steinrück, J. M. Gottfried, N. Lin, *J. Am. Chem. Soc.* **2012**, *134*, 6401–6408.
[80] F. Studener, K. Müller, N. Marets, V. Bulach, M. W. Hosseini, M. Stöhr, *J. Chem. Phys.* **2015**, *142*, 101926.
[81] T. Lin, G. Kuang, X. S. Shang, P. N. Liu, N. Lin, *Chem. Commun.* **2014**, *50*, 15327–15329.

Since nitrogen-metal and transition metal-carbonitrile coordination is highly stabilizing, Schlickum *et al.*[82] used a simple ditopic organic linker, carrying two cyano groups, as molecular brick. In the presence of Co atoms, they formed a size and shape-controlled nanomesh on the Ag (111) surface. The resulting metal-organic honeycomb was thermally robust and its pore size could be tuned by varying the length of the polyphenyldicarbonitrile linker **19** (Figure 16a).

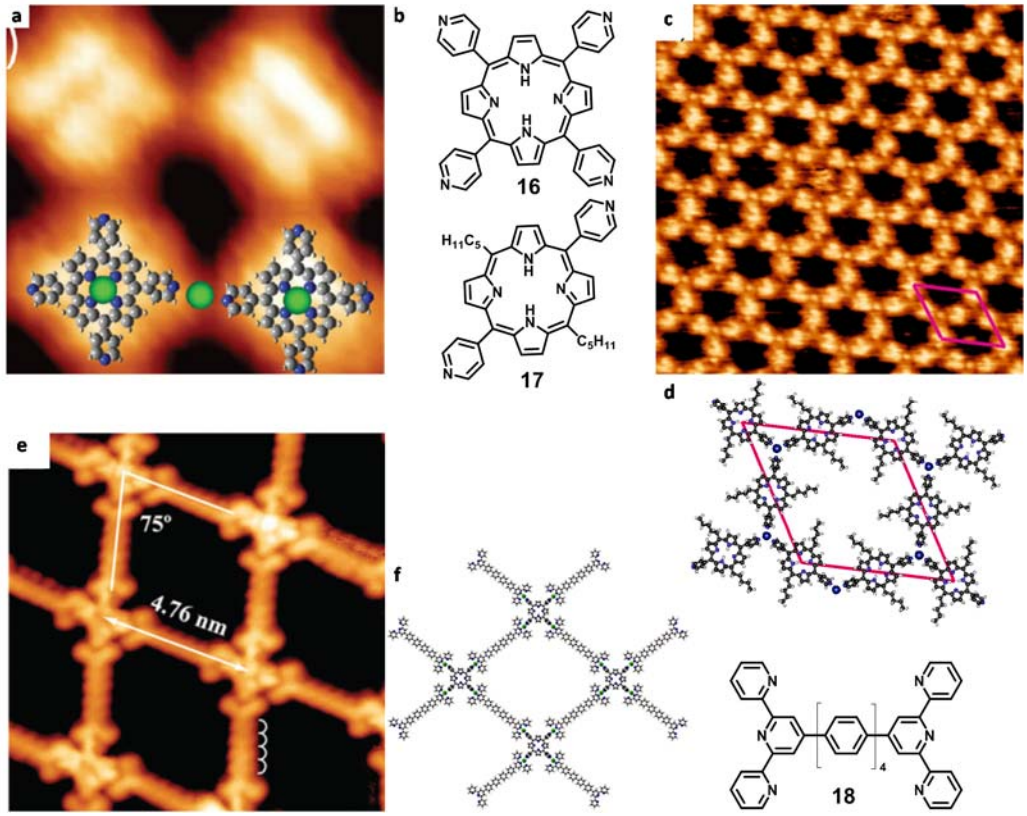


Figure 15. UHV-STM images of metal-coordinated porous networks. (**a**) High-resolution STM image of **16** with Cu on Au (111) at 300 K (4.5 x 5 nm^2, -0.79 V, 0.29 nA). (**b**) Chemical structure of compounds **16** and **17**. (**c**) STM image of the Co-coordinated network of **17** on Ag (111) (20 x 20 nm^2, -1.0 V, 130 pA, rt). The unit cell is indicated in pink. (**d**) Tentative structure model, where one Co atom (blue sphere) coordinates to three pyridyl ligands from **17**. (**e**) High-resolution UHV-STM image (12 x 12 nm^2) of the rhombic network formed by Fe, **16** and **18**; along with (**f**) tentative model and chemical structure of **18**.

Metal adatoms can be thermally generated from the substrate. The term corresponds to a contraction of the formula “adsorbed atom” and refers to the metal that lies on a crystal surface, contrarily to a surface vacancy. Their surface concentration depends on the substrate temperature, as they are often a thermodynamically unfavorable state; and they

[82] U. Schlickum, R. Decker, F. Klappenberger, G. Zoppellaro, S. Klyatskaya, M. Ruben, I. Silanes, A. Arnau, K. Kern, H. Brune, J. V. Barth, *Nano Lett.* **2007**, *7*, 3813–3817.

are frequently key in the study of the self-assembly of organic arrays. It is, for instance, the case of the work published by Björk and co-workers,[83] where a porous metal-organic coordination network (MOCN) of a tetraazapyrene molecule (**20**) was obtained on the Cu (111) substrate *via* the coordination of each N atom to a Cu adatom (Figure 16b).

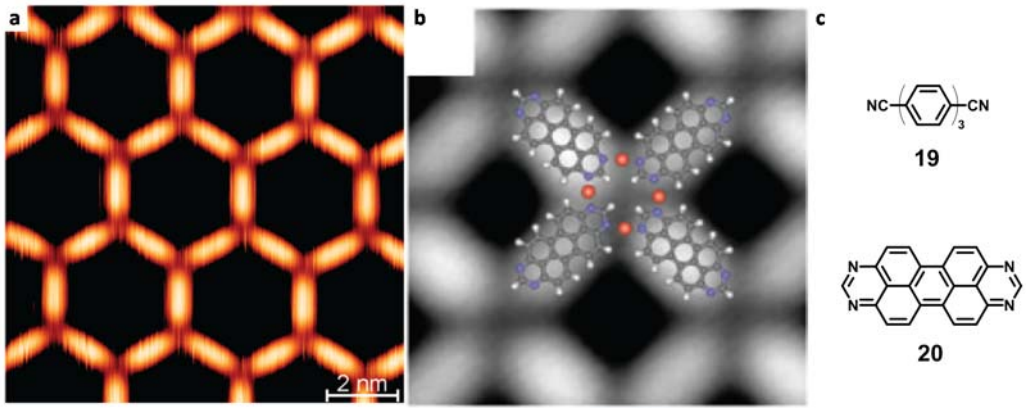


Figure 16. UHV-STM images of 2D-porous networks. (**a**) STM image of the Co-directed assembly of **19** (0.1 nA, 0.9 V). (**b**) **20** on Cu (111) (4 x 4 nm^2, 1 V, 20 pA). Cu adatoms are represented by red spheres. (**c**) Chemical structure of compounds **19** and **20**.

Another functional group vastly used in metal-ligand coordination for the stabilization of porous 2D architectures is the carboxylic moiety. First in 2004, Stepanow *et al.*[84] coordinated Fe to different polytopic organic carboxylate linker molecules, like **21**, on Cu (100) under UHV conditions (Figure 17a). Later, the same group made again use of Fe atoms to fabricate another porous metal-organic architecture on Cu (100), based on terephthalate (**22**) (Figure 17b).[85] Komeda and co-workers[86] used Mn with stillbenedicarboxylic acid (**23**) to observe a square-shaped porous MOCN on the Au (111) surface (Figure 17c).

[83] J. Björk, M. Matena, M. S. Dyer, M. Enache, J. Lobo-Checa, L. H. Gade, T. A. Jung, M. Stöhr, M. Persson, *Phys. Chem. Chem. Phys.* **2010**, *12*, 8815–8821.
[84] S. Stepanow, M. Lingenfelder, A. Dmitriev, H. Spillmann, E. Delvigne, N. Lin, X. Deng, C. Cai, J. V. Barth, K. Kern, *Nature Mat.* **2004**, *3*, 229–233.
[85] A. P. Seitsonen, M. Lingenfelder, H. Spillmann, A. Dmitriev, S. Stepanow, N. Lin, K. Kern, J. V. Barth, *J. Am. Chem. Soc.* **2006**, *128*, 5634–5635.
[86] Y.-F. Zhang, N. Zhu, T. Komeda, *J. Phys. Chem. C* **2007**, *111*, 16946–16950.

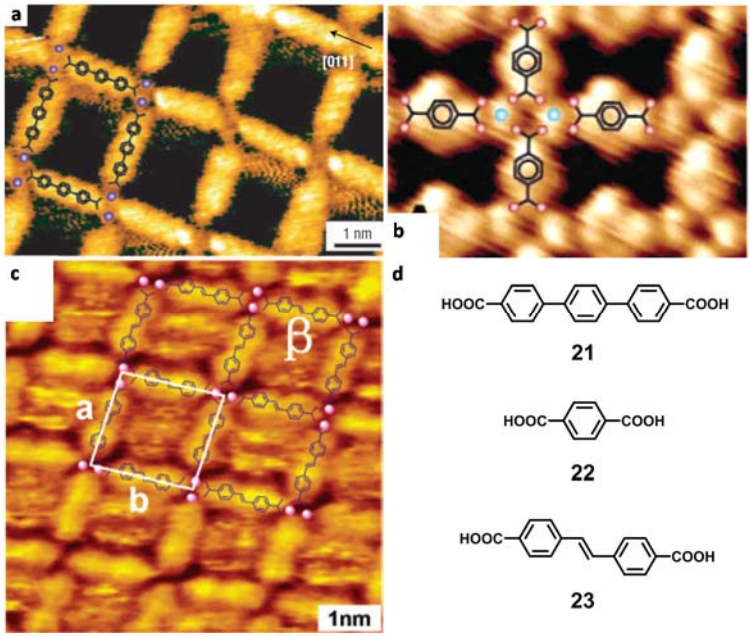


Figure 17. UHV-STM images of porous arrays based on carboxylic acids along with the corresponding molecular models. (**a**) **21**-Fe coordination network on Cu (100). (**b**) Fe-**22** grid on Cu (100) (4 x 3 nm^2). The blue spheres in the models correspond to Fe atoms. (**c**) Mn-**23** network. (**d**) Chemical structure of the compounds **21**-**23**.

More than the strength of the intermolecular interactions, it is their directional nature that allows the formation of the network. Since H-bonding is stronger than van der Waals forces and combines interesting benefits, such as reversibility, directionality, specificity and cooperativity, H-bonded crystals play an important role in the field of supramolecular materials. They are also key in designing 2D porous networks by means of the high level of control over the molecular self-assembly process. From weak H-bonds to multiple and combined classical hydrogen-unions, a vast array of porous structures has been stabilized on the surface, both under UHV and ambient conditions. In 2011, the IUPAC pointed out that a H-bond occurs between X and H as long as X is more electronegative than H and there is evidence of bond formation.[87] Therefore, the interaction between a halogen atom and H is considered as a weak H-bond. Fichou and co-workers[88] demonstrated this by replacing three hydrogens by three Cl-atoms in the bay regions of their triazatrinaphthylene derivative (**24**) (Figure 18a) and locking the configuration of the porous molecular self-assembly on top of the HOPG substrate. Intermolecular CH–Cl halogen-based H-bonds would orient and stabilize the 2D crystal.

[87] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci, D. J. Nesbitt, *Pure Appl. Chem.* **2011**, *83*, 1619–1636.
[88] H. Bertrand, F. Silly, M-P. Teulade-Fichou, L. Tortech, D. Fichou, *Chem. Commun.* **2011**, *47*, 10091–10093.

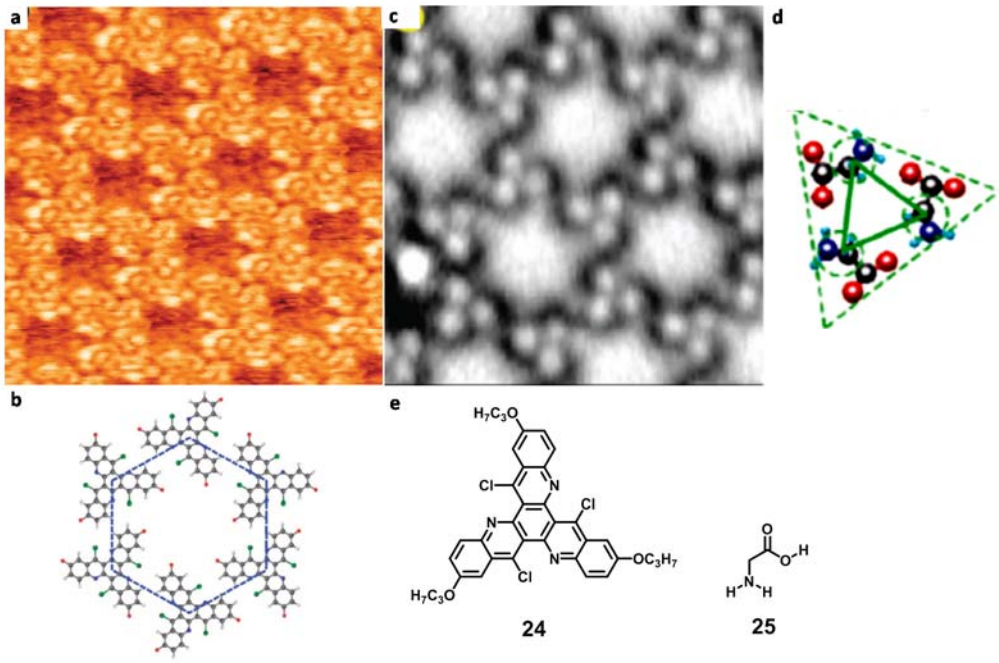


Figure 18. STM images of self-assembled porous networks based on H-bonds. (**a**) High-resolution STM image of **24** at the *n*-tetradecane/HOPG interface (8 x 8 nm^2, 1 V, 10 pA). (**b**) Model of the hexagonal packing of the empty network of **24**. (Cl atoms in green). (**c**) UHV-STM image of the self-assembled network of **25** on Cu (111) (I_t = 0.5 nA, V_s = -0.5 V) along with (**d**) its model. (**e**) Chemical structure of compounds **24** and **25**.

Again, the carboxylic acid moiety is key for pre-programming these H-bonded 2D networks. Shigekawa and co-workers[89] showed, by means of UHV-STM, trimers of the naturally occurring α-amino acid glycine (**25**) on Cu (111) (Figure 18c). More interesting is the H-bond formed between two carboxylic acids, which is the most widely used motif for constructing bidimensional arrays. Among them, Heckl and co-workers showed how trimesic acid (**26**) (Figure 19a,d)[90] and 1,3,5-benzenetribenzoic acid (**27**) (Figure 19g)[91] could form low densely-packed nanoporous surfaces. The trimerization of the carboxylic moieties on surfaces gave the possibility of two polymorphs (*i.e.* other monolayer domain arrangements from the same molecular building blocks) of **26** on the HOPG substrate: chicken-wire (Figure 19a-c) and flower structures (Figure 19d-f) with circular cavities. Compound **27** afforded also a hexagonal lattice exhibiting circular pores (Figure 19g,h).

[89] K. Kanazawa, A. Taninaka, H. Huang, M. Nishimura, S. Yoshida, O. Takeuchi, H. Shigekawa, *Chem. Commun.* **2011**, *47*, 11312–11314.
[90] S. Griessl, M. Lackinger, M. Edelwirth, M. Hietschold, W. M. Heckl, *Single Mol.* **2002**, *3*, 25–31.
[91] L. Kampschulte, M. Lackinger, A.-K. Maier, R. S. K. Kishore, S. Griessl, M. Schmittel, W. M. Heckl, *J. Phys. Chem. B* **2006**, *110*, 10829–10836.

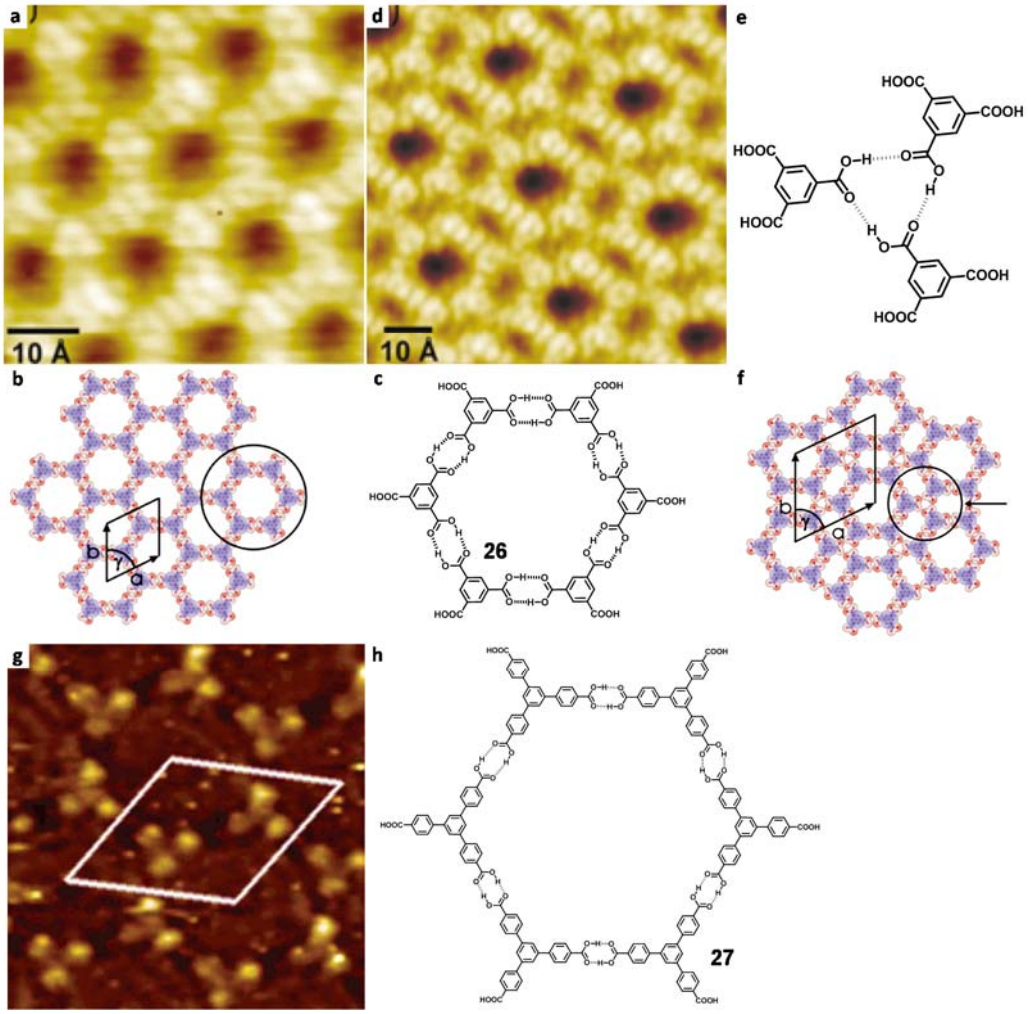


Figure 19. STM images of porous self-assemblies based on carboxylic acids. (**a**) UHV-STM image of **26** on HOPG (5.4 x 5.4 nm^2, -1.4 V, 126 pA). (**b**) Chicken-wire association mode of **26** and (**c**) chemical structure. (**d**) UHV-STM image of **26** on HOPG (8.2 x 8.2 nm^2, -1.4 V, 126 pA). (**e**) Chemical structure of the assembling mode of **26** and (**f**) flower structure. (**g**) **27** at the 1-phenyloctane/HOPG interface. The unit cell is marked in white (1.7 x 3.1 nm^2, 76°). (**h**) Chicken-wire association mode and chemical structure of **27**.

A deeper study in polymorphism was presented by Beton and co-workers,[92] where terphenyl-3,5,3',5'-tetracarboxylic acid (**28**) formed five different 2D porous patterns at the solid-liquid interface (Figure 20a,b). Another example of the utilization of carboxylic acids, this time for the preparation of larger pores, was published in 2010 by Borguet and

[92] M. O. Blunt, J. C. Russell, M. D. Giménez-López, J. P. Garrahan, X. Lin, M. Schroder, N. R. Champness, P.H. Beton, *Science* **2008**, *322*, 1077–1081.

colleagues.[93] They stabilized a square-shaped network of a porphyrin derivative **29** on Au (111) able to host a fifth **29** molecule (Figure 20d).

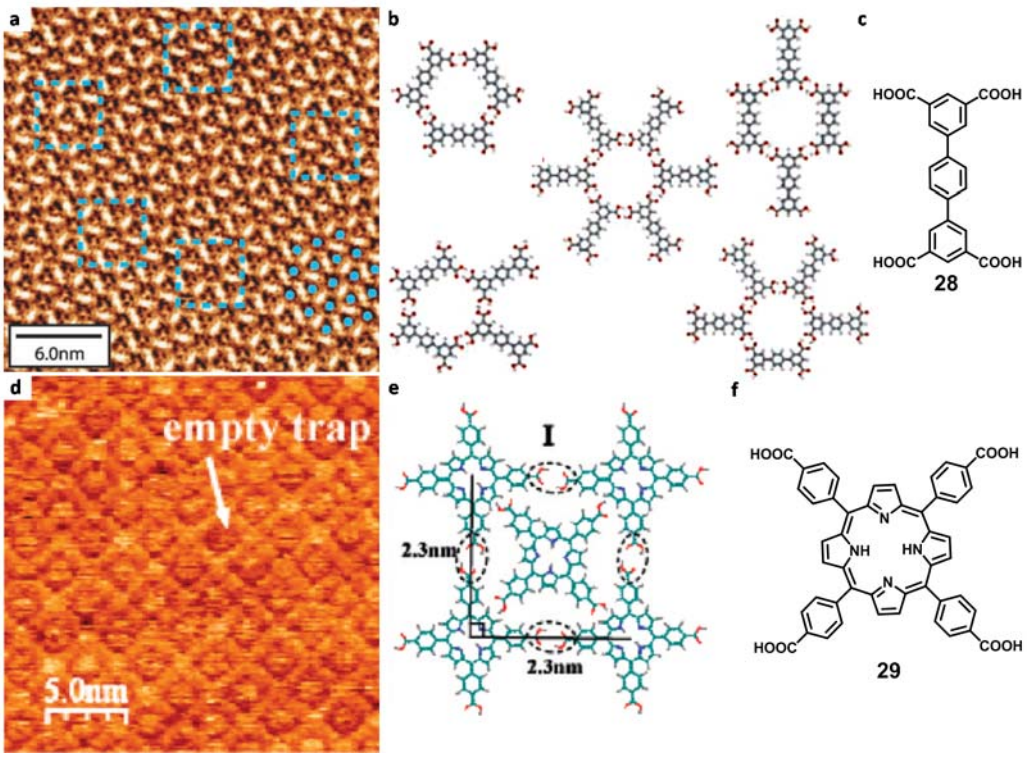


Figure 20. 2D nanoporous assemblies based on carboxylic acids. (**a**) STM image of **28** at the nonanoic acid/HOPG interface (1200 mV, 0.015 nA). (**b**) Different possible association modes for **28**, along with (**c**) its chemical structure. (**d**) STM image of **29** at the $HClO_4$/Au (111) interface (-0.3 V, 0.6 nA (**e**) Model and (**f**) chemical structure of **29**.

Interesting motifs can be constructed by combining two building blocks equipped with a vast array of available H-bonding sites. Fasel and co-workers[94] formed a rectangular porous organic semiconductor nanostructure on Au (111) with 3,4,9,10-perylenetetracarboxylic dianhydride (**30**) and 4,4''-diamino-*p*-terphenyl (**31**) (Figure 21a). Melamine (**32**) appears here as a very useful tool for the construction of self-assembled monolayers, as it presents three *DAD* recognition sites. For instance, Beton and colleagues[95] formed a very stable 2D honeycomb network of large pores with perylene tetra-carboxylic di-imide (**33**) and **32** on Ag/Si(111) (Figure 21c). They were, actually, able to immobilize C_{60}

[93] Q. Yuan, Y. Xing, E. Borguet, *J. Am. Chem. Soc.* **2010**, *132*, 5054–5060.
[94] M. Treier, M.-T. Nguyen, N. V. Richardson, C. Pignedoli, D. Passerone, R. Fasel, *Nano Lett.* **2009**, *9*, 126–131.
[95] J. A. Theobald, N. S. Oxtoby, M. A. Phillips, N. R. Champness, P. H. Beton, *Nature* **2003**, *424*, 1029–1031.

molecules on their organic host architecture as well.[95,96] The group of Samori[97] also studied the formation of a supramolecular bicomponent porous network formed by the ditopic imidic molecule **34**, again with **32**, at the solid-liquid interface (Figure 21d).

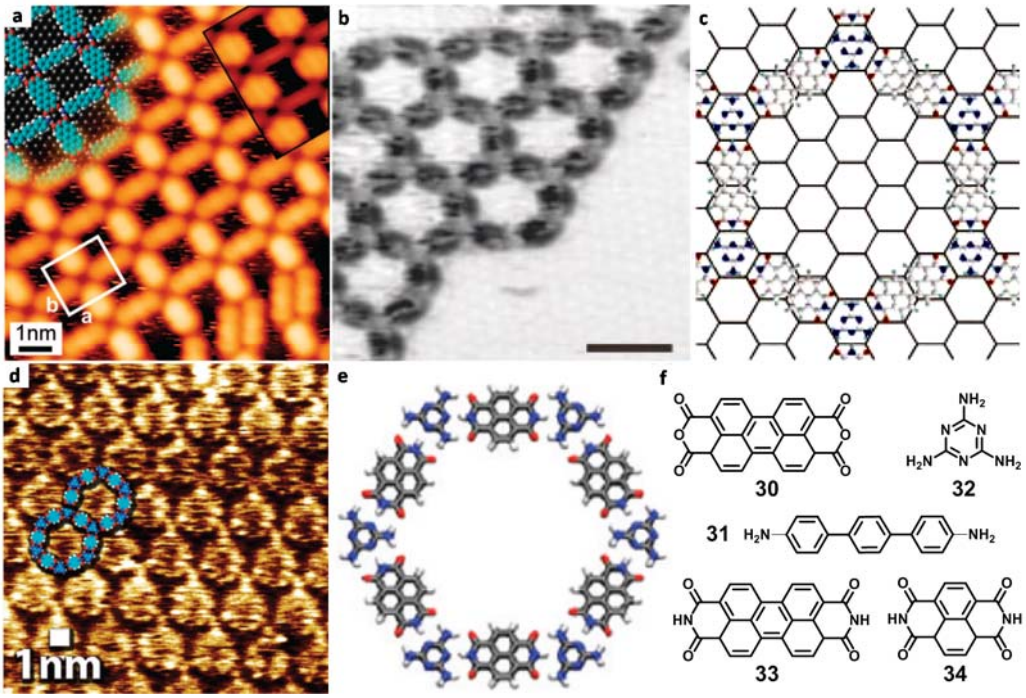


Figure 21. STM images of architectures combining two building blocks. (**a**) High-resolution STM image of **30** and **31** on Au (111) (20 pA, -1.8 V) along with model. (**b**) Inverted STM image of **32** and **33** (0.1 nA, -2 V) on Ag/Si (111) (scale bar 3 nm); along with (**c**) association mode. (**d**) STM image of **32** and **34** at the TCB:DMSO/HOPG interface (5 pA, -150 mV) along with (**e**) model. (**f**) Chemical structures of the compounds **30**-**34**.

2.2. Self-assembled macrocycles onto surfaces.

Controlling the relation between the size of a supramolecular object and its molecular properties is a main target in supramolecular nanostructuration of organic materials. To achieve this goal, the understanding of the self-assembly process has to be mastered. It is important to control the number of molecules that are associating in order to form nanostructures of an extremely precise dimension and with a relatively high monodispersity. In the previous section of this Introduction (**2.1.**), molecules with more than two binding sites were described. Consequently, their self-assembly on the surface

[96] L. M. A. Perdigao, E. W. Perkins, J. Ma, P. A. Staniec, B. L. Rogers, N. R. Champness, P. H. Beton, *J. Phys. Chem. B* **2006**, *110*, 12539–12542.
[97] C.-A. Palma, J. Bjork, M. Bonine, M. S. Dyer, A. Llanes-Pallas, D. Bonifazi, M. Persson, P. Samori, *J. Am. Chem. Soc.* **2009**, *131*, 13062–13071.

formed, inevitably, a 2D network. This section describes planar monocyclic systems where the discrete macrocycle constitutes the primary repeating unit. This entity then forms a network by means of secondary interactions that can be different from the H-bonds that hold the cycle together. As we argued formerly, targeting macrocycles is a crafty approach for this purpose. Compared to the previous presented examples in solution (section **1.2.** of this Introduction), most of these discrete macrocylic systems are bound by relatively weak H-bonding interactions and would be too labile to afford the same structures in solution. The intramolecular event that shifts the equilibria toward cyclization (chelate effect) is only relevant at low concentrations. When the molecules are concentrated in the crystal or onto a substrate, intra- and intermolecular binding events are compensated. Therefore, nothing can assure that closed H-bonded species are going to dominate over open oligomers onto surfaces, even if the molecule is preorganized for ring closure. Macrocycles are either observed at low monolayer coverages or because the overall network is more stable when the monomer assembles in its cyclic form. Cyclic systems would only be formed as the result of a subtle interplay between molecule-substrate, molecule-molecule and molecule-solvent (if any) interactions. Then, a network would be formed by repeating this macrocyclic unit. In these next paragraphs, a few examples of the literature concerning discrete cyclic assemblies, stabilized on the surface through distinct interactions, are presented.

Dipole-dipole interactions occur when the electronegative portion of one molecule is attracted to the electropositive portion of a second molecule. If one of the participating atoms is hydrogen, this interaction can also be seen as a weak and single H-bond. Using low temperature STM, Mashiko and co-workers[98] showed that cyano-substituted porphyrins, like **35**, adsorbed on a gold surface, could form cyclic trimers and tetramers endowed with a small central pore (Figure 22). The estimated length of the CH···NC bond was around 2.6 Å and 2.5 Å, suggesting weak H-bonding interactions.

[98] T. Yokoyama, S. Yokoyama, Y. Okuno, S. Mashiko, *Nature* **2001**, *413*, 619–621.

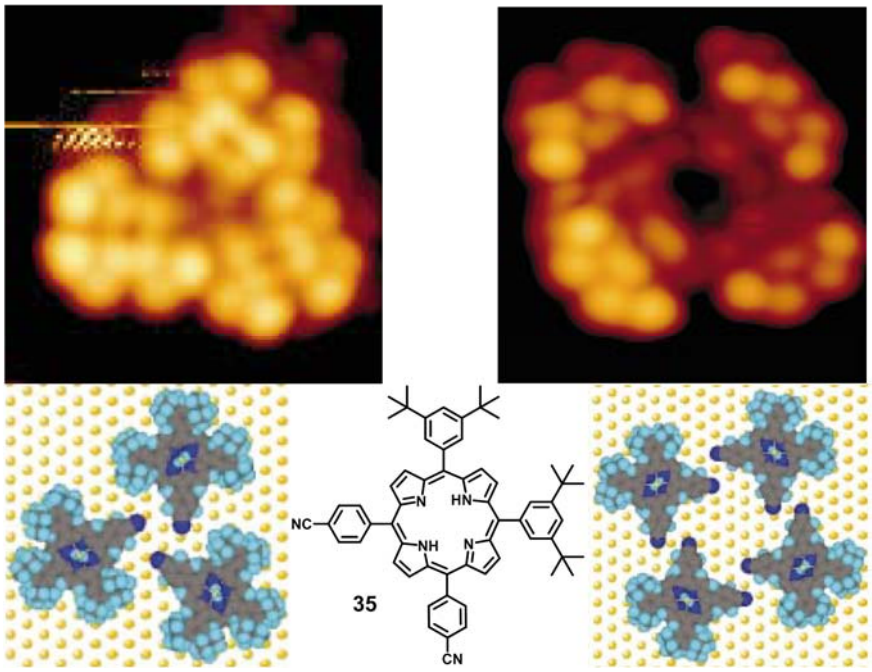


Figure 22. STM images at 63 K of the supramolecular aggregation of compound **35** on Au (111), their corresponding molecular models and the chemical structure of **35**.

Later, the group of Diederich[99] was able to self-assemble on the Cu (111) surface a larger Zn^{II} porphyrin derivative **36** into discrete macrocyclic-polymorphs, growing in size. At sub-monolayer coverage, the *cis*-porphyrin **36** formed cyclic trimers, tetramers, pentamers and hexamers held together by CN···H-C(sp^2) H-bonding (Figure 23), involving the positively polarized hydrogen *ortho* to the CN group of the neighboring molecule. Also, antiparallel CN···CN dipolar interactions helped in the stabilization of the oligomeric macrocycles. The authors argued that the very strong adsorbate/Cu (111) substrate interactions, due to the extended π–system of the porphyrins, compensated the loss in antiparallel CN···CN forces and were the origin of the diversity of the obtained porous macrocycles, compared to the earlier work of Mashiko.[98] On top of it, this group was also able to form related porous layers, again based on cyano-substituted porphyrins, and template the complexation and organization of fullerene nanoarrays on silver.[100]

[99] L.-A. Fendt, M. Stöhr, N. Wintjes, M. Enache, T. A. Jung, F. Diederich, *Chem. Eur. J.* **2009**, *15*, 11139–11150.
[100] a) D. Bonifazi, H. Spillmann, A. Kiebele, M. Wild, P. Seiler, F. Cheng, H.-J. Güntherodt, T. Jung, F. Diederich, *Angew. Chem. Int. Ed.* **2004**, *43*, 4759–4763; b) H. Spillmann, A. Kiebele, T. A. Jung, D. Bonifazi, F. Cheng, F. Diederich, *Adv. Mater.* **2006**, *18*, 275–279.

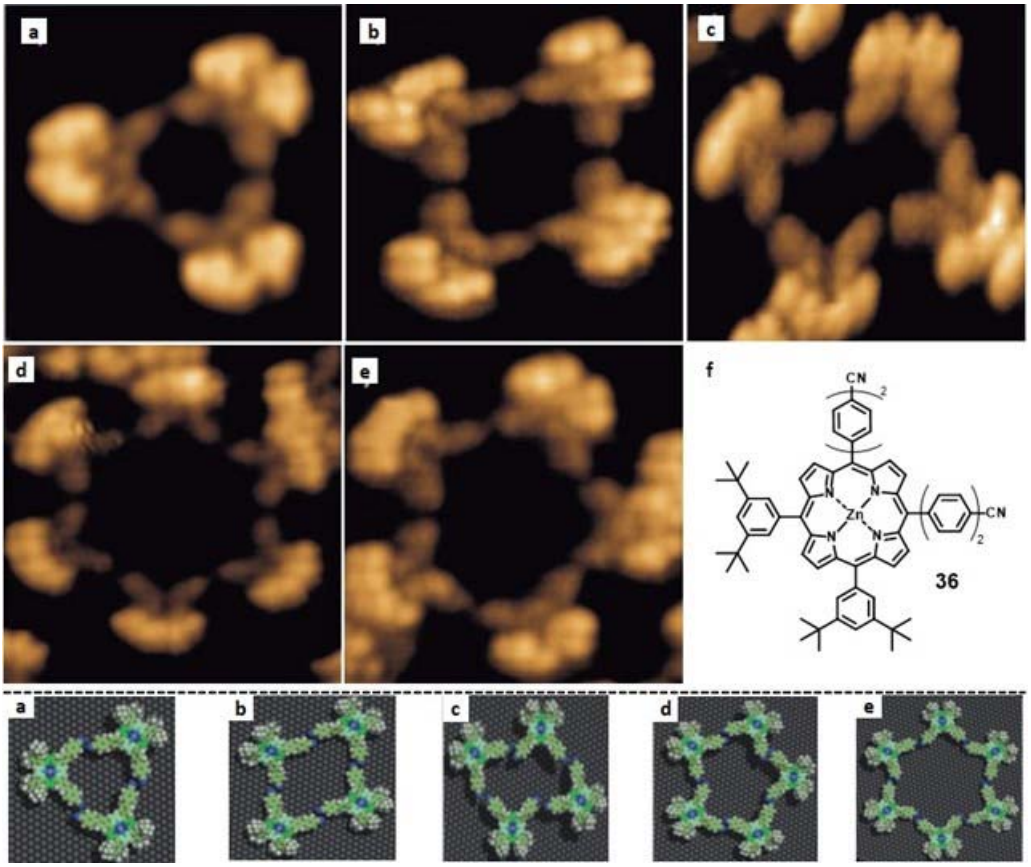


Figure 23. STM images of porphyrin **36** on Cu (111) with corresponding molecular models and chemical structure of **36**.

These H···N≡C interactions featured by associating cyanophenyl rings are substantially weaker than classic H-bonds involving more acidic O-H or N-H donors. One more time, the motif created by two carboxylic acids has inspired many researchers to design well-defined macrocyclic arrangements on the surface. The benzene ring solely *meta*-substituted with two carboxylic acid substituents (isophthalic acid, **37**) do not form a porous nanostructure and prefer a densely-packed network with zigzag linear chains. However, there are ways to circumvent this and promote cycle formation. For instance, Hecht, Grill and colleagues[101] made use of peripheral azobenzene derivatives (**38**) equipped with –COOH groups to direct the supramolecular self-assembly of highly stable discrete hexameric rosettes (Figure 24a,b). In the rosette, each molecule was connected to both of its neighbors and formed the maximum number of non-distorted carboxylic acid dimer motifs. Therefore, the binding energy *per* molecule was maximized. They measured a distance of 3.6 ±0.7 Å between oxygen atoms, so the pore diameter was around 1.7 nm. On the other hand, De Feyter and

[101] S. Selvanathan, M. V. Peters, J. Schwarz, S. Hecht, L. Grill, *Appl. Phys. A* **2008**, *93*, 247–252.

co-workers[102] proved how coronene (**cor**) could act as a molecular template directing the assembly of isophthalic acid **37** and afford a heteromolecular cluster (Figure 24d,e). Six molecules were held together by 12 H-bonds, leaving a cavity of ~1 nm in diameter, where **cor** fitted perfectly and further stabilized this 2D crystal *via* van der Waals interactions with a force of -54 kJ/mol, according to their molecular mechanics calculations. Larger pores able to host a Cu^{II}-phthalocyanine were also obtained by Bai and colleagues[103] with the triple-armed amphiphile **39** (Figure 24f,g).

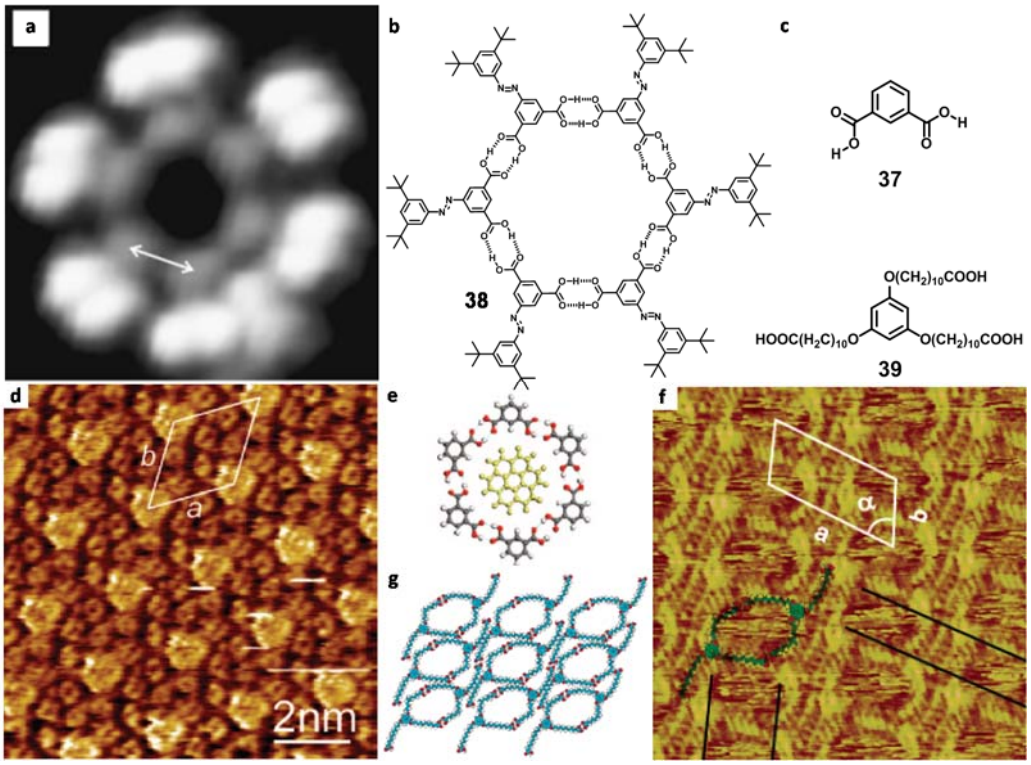


Figure 24. 2D nanoporous self-assembled systems based on carboxylic acids. (**a**) STM image of **38** (5.2 x 5.2 nm^2; I = 0.11 nA; V = 1 V) on Au (111) and (**b**) structure. (**c**) Chemical structure of compounds **37** and **39**. (**d**) STM image at the 1-octanoic acid/HOPG interface of the self-assembled pattern of **37** and **cor** (I_{set} = 530 pA, V_{set} = -1.00 V), along with (**e**) the corresponding molecular model and chemical structure. (**f**) STM image of the network formed by **39** on HOPG along with (**g**) model.

Other molecular families carrying O···H bonding moieties can be used. Mamdouh *et al.*[104] imaged at the solid-liquid interface porous discrete cyclic tetramers from an alkylated monodendron **40** (Figure 25a). The formation of tetramers was due to H-bonding between the carboxylic group of one monodendron and the hydroxyl group of another one. The alkyl

[102] S. Lei, M. Surin, K. Tahara, J. Adisoejoso, R. Lazzaroni, Y. Tobe, S. De Feyter, *Nano Lett.* **2008**, *8*, 2541–2546.
[103] J. Lu, S.-B. Lei, Q.-D. Zeng, S.-Z. Kang, C. Wang, L.-J. Wan, C.-L. Bai, *J. Phys. Chem. B* **2004**, *108*, 5161–5165.
[104] W. Mamdouh, H. Uji-i, A. E. Dulcey, V. Percec, S. De Feyter, F. C. De Schryver, *Langmuir* **2004**, *20*, 7678–7685.

chains helped in reducing the mobility of the molecules when physisorbed on graphite. Stöhr *et al.*[105] also used classic H-bonds, involving in this case N-H moieties. They stabilized discrete trimers of the perylene derivative **41** on the Cu (111) surface by intermolecular resonance-assisted H-bonding. Upon annealing a sample with a coverage of 0.85 mL **41** at 300 °C, trimeric structures in the form of rings were found (Figure 25d).

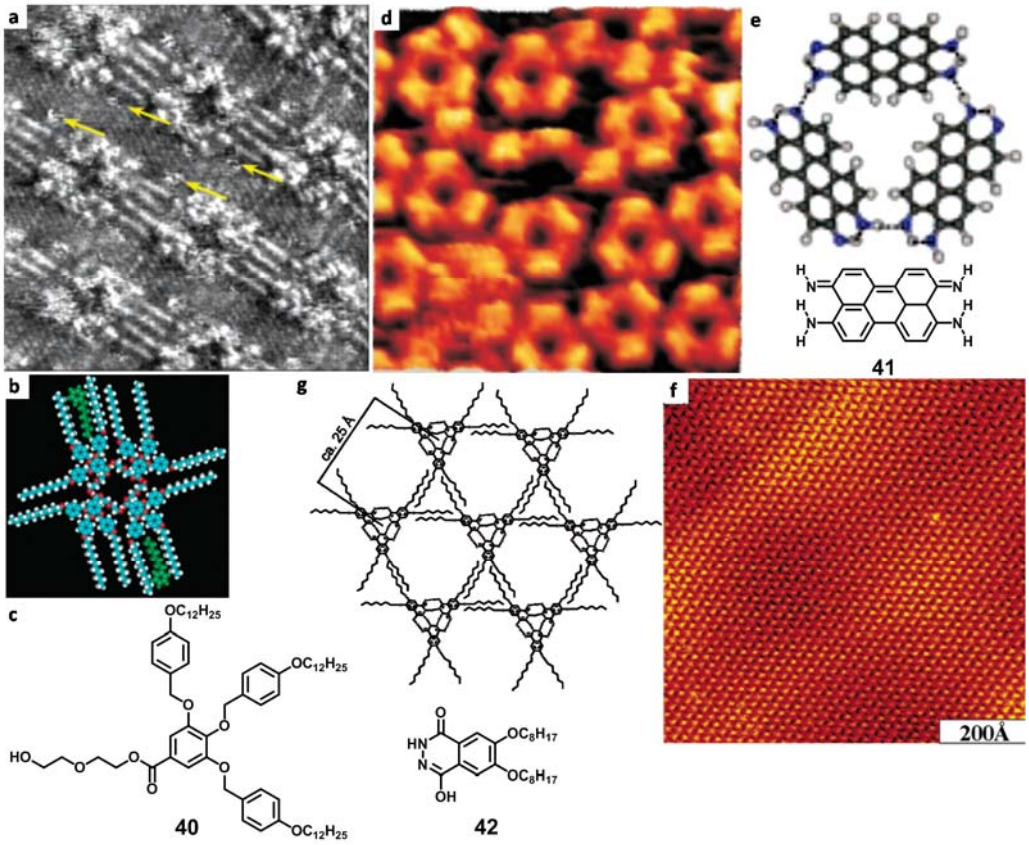


Figure 25. STM images of porous 2D networks. (**a**) STM image of **40** at the 1-phenyloctane/HOPG interface (I_t = 0.5 nA, V_{bias} = -0.518 V). (**b**) Tentative molecular model and (**c**) chemical structure of **40**. (**d**) STM image of **41** on Cu (111) along with (**e**) association mode and chemical structure of **41**. (**f**) STM image of **42** at the 1-chloronaphtalene/HOPG interface (I_t = 27 pA, V_s = 251 mV). (**g**) Corresponding model and chemical structure of **42**.

Some research groups have combined N– and O– H-bonding in order to get more interesting self-assembled macrocycles. For example, Ziener and co-workers[106] obtained trimers from an interplay of both H-bonds and van der Waals interactions of lactim/lactam moieties (Figure 25f). The synthesized phthalhydrazide **42** undergoes tautomerism between three forms and discrete cyclic units were formed if neighboring molecules

[105] M. Stöhr, M. Wahl, C. H. Galka, T. Riehm, T. A. Jung, L. H. Gade, *Angew. Chem. Int. Ed.* **2005**, *44*, 7394–7398.
[106] A. Mourran, U. Ziener, M. Möller, M. Suarez, J. M. Lehn, *Langmuir* **2006**, *22*, 7579–7586.

stablished double H-bonds in both sides of the molecule, through their two N–C–O H-bonding moieties. The trimer was formed by six H-bonds in the lactim/lactam tautomeric form. Matzger and colleagues[107] formed hexamers based on an amphiphilic amide **43**, further stabilized by van der Waals interactions between the long alkyl chains (Figure 26a,b). The authors described how, depending on the conformation of the amide, 3- and 6-fold symmetric H-bonding patterns can be seen. In the 3-fold rotation axis, a typical arrangement of three dimers, united by a single H-bond each, took place. Hexameric rings involving 12 H-bonding donors and 6 acceptors existed, as well, at the 6-fold axis. In this case, the two oxygen lone pairs made H-unions with two hydrogen atoms in an amide. An amide was also used as H-bonding carrier by the group of Cousty,[108] who was able to stabilize hexamers of 2-pyrrolidone **44** on Au (111) (Figure 26c). A periodic arrangement of hexamers was observed (unit cell parameters: a = 1.33 ±0.05 nm, b = 1.33 ±0.05 nm and γ = 120 ±1°) corresponding to cyclic pyrrolidone funnel-like hexamers. A model was proposed where each molecule stablished one H-bond with each one of their two nearest neighbors and the plane of each molecule was tilted with respect to the gold substrate. The distance measured between neighboring residues was 0.5 nm and fitted the funnel-like arrangement. Theoretical calculations showed that this macrocycle was only slightly more stable than the pyrrolidone dimer. However, the hexameric shape fitted better the Au (111) lattice and the molecule-substrate interactions were, therefore, stronger than for a dimer, explaining the absence of polymorphism.

H-bonded assemblies in which the individual components carry opposite charges have higher stabilities than assemblies consisting of neutral components and can be seen as ion-pair-reinforced motifs.[37a] Recently, Kunkel *et al.*[109] prepared strongly dipolar *p*-benzoquinonemonoimines to study the competition between chemical interactions and substrate effects in the self-assembly on metal surfaces. When sublimating the butyl zwitterionic molecule **45** and making the deposition on Au (111), ring-shaped tetramers stabilized by four H-bonds were obtained (Figure 26e). In this case the geometrical hindrance of the alkyl substituents was an important factor to limit the size of the cluster. For instance, in the case of the parent *p*-benzoquinone-monoimine, with no alkyl substituent and therefore smaller constraints, larger rings consisting of six molecules were most frequently observed. For this bigger self-assembled macrocycle, the pore was large enough to host a seventh **45** molecule.

[107] S. Ahn, C. N. Morrison, A. J. Matzger, *J. Am. Chem. Soc.* **2009**, *131*, 7946–7947.
[108] Z. X. Xie, J. Charlier, J. Cousty, *Surf. Sci.* **2000**, *448*, 201–211.
[109] D. A. Kunkel, J. Hooper, S. Simpson, D. P. Miller, L. Routaboul, P. Braunstein, B. Doudin, S. Beniwal, P. Dowben, R. Skomski, E. Zurek, A. Enders, *J. Chem. Phys.* **2015**, *142*, 101921.

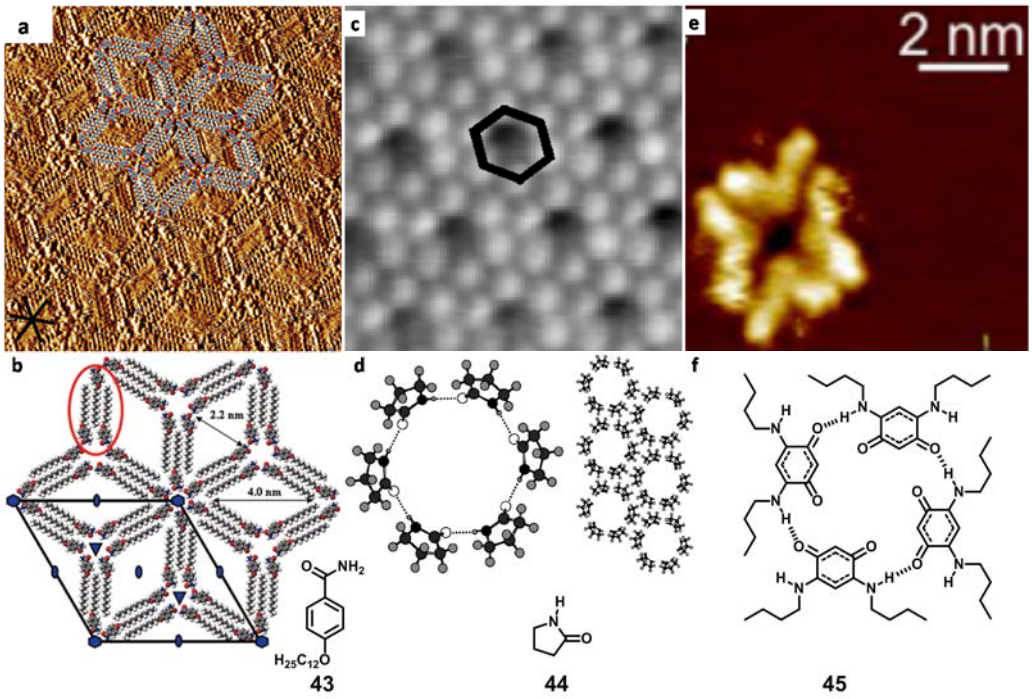


Figure 26. STM images of macrocycles based on the amide motif. (**a**) STM image of **43** at the 1-phenyloctane/HOPG interface (I_t = 300 pA, V_s = 0.80 V). (**b**) Computed model and chemical structure of **43**. (**b**) STM image of **44** on Au (111) and (**c**) model and structure. (**e**) STM image of **45** on Ag (111). (**f**) Chemical structure of **45** and association model.

Successive H-bond motifs allow the stabilization of more complex macrocycles. When increasing the number of H-bonds to three parallel ones in a *ADA–DAD* pair, the association energy rises to very high values. With its three *DAD* faces, melamine **32** is, once again, a very useful synthon for creating discrete porous architectures. In 2009, Walch *et al.*[110] were able to obtain stable monolayers of melamine as six-membered rings, stabilized by 12 internal N···H-N bonds at the liquid/graphite interface. In these systems, fatty acids, such as nonanoic acid, were used as solvent. It is interesting to see that nonanoic acid not only served as a medium for the solute molecules, but it also stabilized and isolated the macrocycles on the surface. Twelve solvent molecules arranged radially around molecule **32**, stablishing a total of 36 H-bonds and allowing the formation of crystalline hexameric cycles (Figure 27a,b). Melamine formed very tight complexes with carboxylic acids as a result of proton transfer.

Using a related H-bonding motif, the diaminotriazine, Jonkheijm *et al.*[111] stabilized at the solid-liquid interface hexameric π-conjugated rosettes from molecule **46**, which bore a

[110] H. Walch, A-K. Maier, W. M. Heckl, M. Lackinger, *J. Phys. Chem. C* **2009**, *113*, 1014–1019.
[111] P. Jonkheijm, A. Miura, M. Zdanowska, F. J. M. Hoeben, S. De Feyter, A. P. H. J. Schenning, F. C. De Schryver, E. W. Meijer, *Angew. Chem. Int. Ed.* **2004**, *43*, 74–78.

oligo(*p*-phenylenevinylene) (OPV) fragment (Figure 27c,d). Later, Hoeben *et al.*[112] saw how the true *ADA–DAD* motif of melanine could stabilize other self-assembled macrocycles at the 1-phenyloctane/HOPG interface. By applying, onto the freshly cleaved HOPG substrate, a drop of the V-shaped-molecule **47** solution, trimers and tetramers were observed (Figure 27e,f). In the first example,[111] each molecule stablished four H-bonds for stabilizing the six-membered rosette (~7 Å pore diameter) with six hydrogen *D–A* pairs. In the second one,[110] the bifunctional molecules with two H-bonding arrays available for complementary binding allowed the formation of smaller discrete macrocycles.

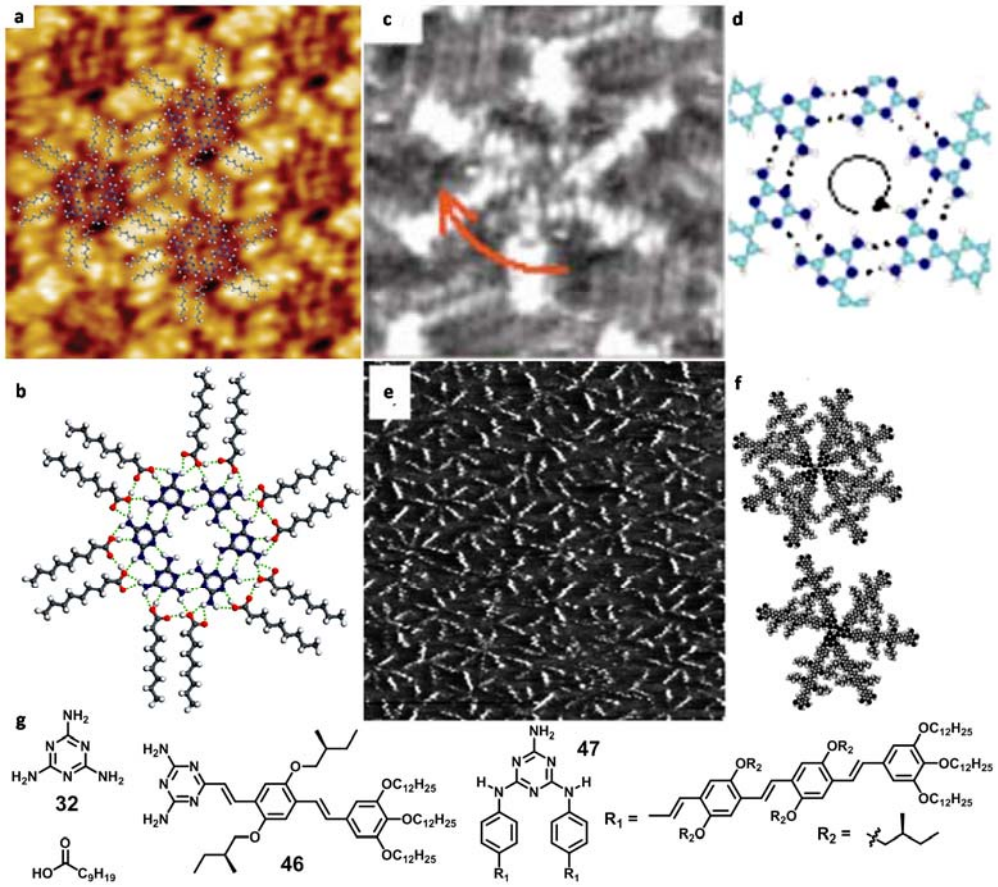


Figure 27. STM images of self-assembled macrocycles based on melamine. (**a**) STM image of **32** at the nonanoic acid/HOPG interface (I_t = 110 pA, V_s = 1.0 V). (**b**) Molecular model. (**c**) STM image of **46** at the 1-phenyloctane/HOPG interface (I_t = 0.50 nA, V_s = -0.50 V) along with (**d**) model. (**e**) STM image of **47** at the 1-phenyloctane/HOPG interface (I_t = 0.48 nA, V_s = -0.486 V) along with (**f**) corresponding association modes. (**g**) Chemical structures of the considered compounds.

[112] F. J. M. Hoeben, J. Zhang, C. C. Lee, M. J. Pouderoijen, M. Wolffs, F. Würthner, A. P. H. J. Schenning, E. W. Meijer, S. De Feyter, *Chem. Eur. J.* **2008**, *14*, 8579–8589.

2D porous networks serve as template for the controlled self-assembly of single-molecule-based devices. We have seen many examples of alternatives to covalently built cavities that generate hosts *in situ* at the surface during the adsorption process, by means of the supramolecular assembly of smaller complementary components. The incorporation, at this point, of suitable guest molecules to the porous host systems appears as the next step in constructing appropriate materials for nanoscale functional devices. Due to their electrochemical and photophysical properties,[113] fullerene and its derivatives have been broadly investigated as building blocks to prepare self-assembled electronic materials. Among them, Pan *et al*.[114] were able to construct an ordered array of C_{60} molecules placed within flexible vase-shaped calixarene molecules, at the electrolyte/Au (111) interface. A key feature in the designed host-guest systems should be the rigidity of the molecular network involving permanent porosities. Therefore, robust self-assembled macrocycles on the surface have to be designed in order to obtain such appealing functional nanoscale architectures.

2.3. The use of nucleobases to guide 2D assembly.

On account of the characteristics of the work presented within this Thesis, we thought it would be appropriate to include a section concerning the self-assembly of nucleobases and their analogues in two dimensions. Hereafter are collected a few studies concerning U, T, C, *iso*cytosine (iC), A and G derivatives among others, both on metallic substrates and HOPG. Also, some mixtures of DNA-based derivatives are considered, in view of addressing the issue of Watson-Crick complementarity.

In 1997, Nakagawa *et al.*[115] studied the self-assembly of U molecules on Cu (111). They saw how at room temperature the nucleobase diffused over the surface and could not be imaged due to the weak adsorbate/substrate interactions. However at 70 K, U formed trimers at low coverage, and then a hexagonal superstructure when increasing the coverage. Their results showed that the molecular plane of the molecules laid parallel to the Cu (111) surface and that the molecules were connected *via* anisotropic H-bonding. They suggested a mechanism of formation where firstly U trimers were formed and then assembled into larger islands, maintaining their structure. The group of Besenbacher analyzed in depth the self-assembly of T molecules on Au (111).[116] Initially, at low coverages, strong multiple H-bonds steered the formation of a kinetically hierarchical architecture of 1D T filaments (Figure 28a). Then, these self-assembled laterally into 2D

[113] a) L. Echegoyen, L. E. Echegoyen, *Acc. Chem. Res.* **1998**, *31*, 593–601; b) D. M. Guldi, M. Prato, *Acc. Chem. Res.* **2000**, *33*, 695–703.
[114] G.-B. Pan, J.-M. Liu, H.-M. Zhang, L.-J. Wan, Q.-Y. Zheng, C.-L. Bai, *Angew. Chem. Int. Ed.* **2003**, *42*, 2747–2751.
[115] T. Nakagawa, H. Tanaka, T. Kawai, *Surf. Sci.* **1997**, *370*, 144–148.
[116] W. Xu, R. E. A. Kelly, R. Otero, M. Schöck, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2007**, *3*, 2011–2014.

islands, stabilized by weak van der Waals interactions (Figure 28b,c). Otero *et al*.[117] showed how C formed a random network on Au (111) under UHV conditions. 1D zigzag filaments interconnected by five- and sixfold rings were formed. C–C dimers can be formed through the union between binding-sites 5 or 7 and H-bonding can also take place between one binding-site type and the other (Figure 28d). Ciesielski *et al*.[118] prepared an iC derivative **48** bearing two different faces: one with an *AA* H-bonding pattern and the other presenting a *DD* motif. In that way, they crystallized a heterohexameric macrocyle on the HOPG/TCB interface, further stabilized by intermolecular van der Waals interactions between the six-membered assemblies (Figure 28e,f). The measured unit cell parameters led to an area of 6.54 ±0.27 nm^2 containing six iC derivatives, a single molecule occupying an area of 1.09 ±0.05 nm^2 and the pore size being around 6 Å. On the other hand, adenine self-assembled into a well-ordered hexagonal nanostructure onto Au (111) (Figure 28g).[119]

[117] R. Otero, M. Lukas, R. E. A. Kelly, W. Xu, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Science* **2008**, *319*, 312–315.
[118] A. Ciesielski, S. Colella, L. Zalewski, B. Bruchmann, P. Samori, *Cryts. Eng. Comm.* **2011**, *13*, 5535–5537.
[119] R. E. A. Kelly, W. Xu, M. Lukas, R. Otero, M. Mura, Y.-J. Lee, E. Laegsgaard, I. Stensgaard, L. N. Kantorovich, F. Besenbacher, *Small* **2008**, *4*, 1494–1500.

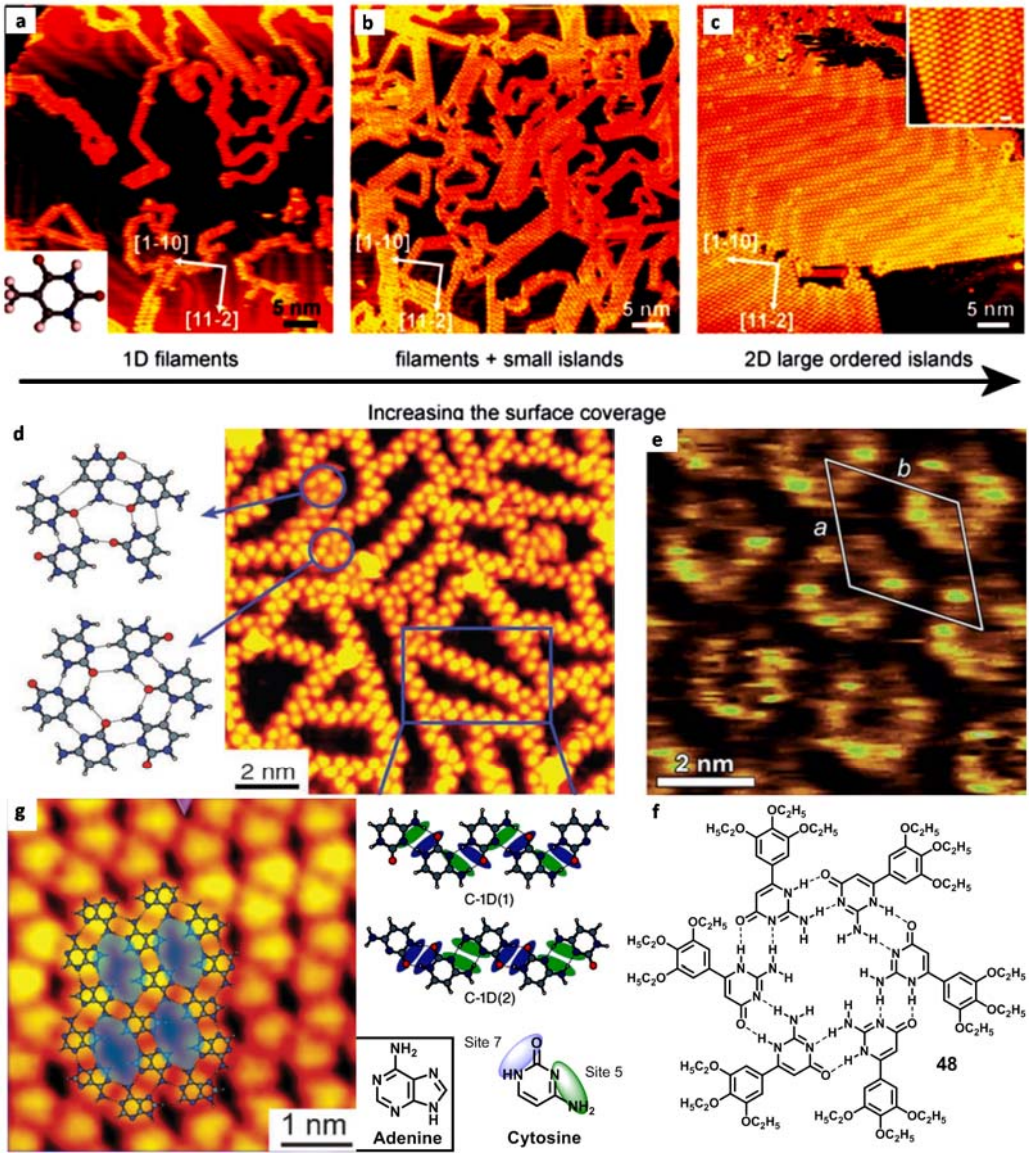


Figure 28. Nucleobases 2D-self-assemblies. (**a**) UHV-STM image of several 1D T filaments with random growth directions on Au (111) (I_t = -0.23 nA, V_t = -1051 mV). The lower inset shows the chemical structure of the thymine molecule. (**b**) UHV-STM image containing 1D T filaments and small islands (I_t = -0.61 nA, V_t = -1486 mV). (**c**) UHV-STM image of extended 2D ordered T islands observed when the surface coverage is increased to ~0.3–0.4 mL (I_t = -0.65 nA, V_t = -2102 mV). The inset shows a zoomed STM image of the 2D T island structure with high molecular resolution. (**d**) UHV-STM image of cytosine on Au (111) (I_t = -0.7 nA, V_t = -1250 mV) along with models. (**e**) STM image of **48** at the TCB/HOPG interface (I_t = 17 pA, V_s = 450 mV) and (**f**) model. (**g**) UHV-STM image of adenine on Au (111) (I_t = 0.5 nA, V_t = 1250 mV) along with model.

The self-assembly of DNA bases has played a significant role in constructing hierarchical nanostructures, the most important ones incorporating guanine derivatives. Both hydrophilic and hydrophobic G-derivatives hold attractive applicability, for instance in electronics, taking advantage of its low oxidation potential. G-derivatives show strong self-association phenomena, mainly because they can bind both through their Watson-Crick and Hoogsteen faces. G-derivatives can form high order structures also through their metal-chelation ability. These structures include ribbons, tetramers, octamers, quadruplexes and more complex architectures, depending on the conditions. Gottarelli *et al*.[120] explored the self-assembly of a lipophilic guanosine at the TCB/HOPG interface and saw how the molecular arrangement fitted a ribbon-like structure, also present in crystals, as seen by X-ray diffraction.[121] The bases laid flat on the plane of graphite and self-assembled by stablishing H-bonds between *N*-1H···*N*-7 and *N*-2H···O6. In this way, guanosine derivatives self-associated through H-bonding both in the Watson-Crick and Hoogsteen edges, forming an angle of 90º between nucleobases.

The G quartet (G_4) was first identified in 1962 as the basis for the aggregation of 5'-guanosine monophosphate[122] and it can be seen as an ideal scaffold to construct functional assemblies.[56,123] Like in the previous work on guanosine derivatives,[120] Ciesielski *et al*.[124] observed how their G-derivative **49** self-assembled onto HOPG as ribbons, unless it formed a cation-templated assembly with K^+ when adding 100 eq of potassium picrate (Figure 29a,b). The cation played a crucial role in stabilizing the G tetramer through ion-dipole interactions with the inner carbonyl moieties. The unit cell contained four **49** molecules and led to an area of 5.21 ±1.35 nm^2. The authors attributed the high-contrast circular feature to the K^+ ions not being located in the same plane as the G_4, as it was also the case in solution.[125] The observed packing motif differed also from the one observed in UHV without templating ions (Figure 29c).[126] Another interesting approach for the stabilization of the G_4 on the surface was published by González-Rodríguez *et al.* that same year.[127] They capped a guanosine derivative with an OPV oligomer and observed how compound **50** formed a monolayer of individual quartets onto HOPG, without any cation-templated stabilization of the assembly (Figure 29d). The authors argued that the bulky ribose group blocked the NH_2 proton and the formation of G-ribbons was prevented because of steric hindrance, resulting in the stabilization of the circular tetramers by H-bonding through the Hoogsteen edges.

[120] G. Gottarelli, S. Masiero, E. Mezzina, S. Pieraccini, J. P. Rabe, P. Samori, G. P. Spada, *Chem. Eur. J.* **2000**, *6*, 3242–3248.
[121] U. Thewalt, C. E. Bugg, R. E. Marsh, *Acta Crystallogr. Sect. B* **1970**, *26*, 1089–1101.
[122] M. Gellert, M. N. Lipsett, D. R. Davies, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018.
[123] a) S. Lena, S. Masiero, S. Pieraccini, G. P. Spada, *Chem. Eur. J.* **2009**, *15*, 7792–7806; b) S. Masiero, G. Gottarelli, S. Pieraccini, *Chem. Comm.* **2000**, *20*, 1995–1996. c) S. Martic, X. Liu, S. Wang, G. Wu, *Chem. Eur. J.* **2008**, *14*, 1196–1204; d) G.P. Spada, S. Lena, S. Masiero, S. Pieraccini, M. Surin, P. Samori, *Adv. Mater.* **2008**, *20*, 2433–2439.
[124] A. Ciesielski, S. Lena, S. Masiero, G. P. Spada, P. Samori, *Angew. Chem. Int. Ed.* **2010**, *49*, 1963–1966.
[125] A. L. Marlow, E. Mezzina, G. P. Spada, S. Masiero, J. T. Davis, G. Gottarelli, *J. Org. Chem.* **1999**, *64*, 5116–5123.
[126] R. Otero, M. Schock, L. M. Molina, E. Laegsgaard, I. Stensgaard, B. Hammer, F. Besenbacher, *Angew. Chem. Int. Ed.* **2005**, *44*, 2270–2275.
[127] D. González-Rodríguez, P. G. A. Janssen, R. Martín-Rapún, I. De Cat, S. De Feyter, A. P. H. J. Schenning, E. W. Meijer, *J. Am. Chem. Soc.* **2010**, *132*, 4710–4719.

Also, the OPV moiety granted compound **50** with additional stability on HOPG, compared to Gottarelli's lipophilic guanosine.[120]

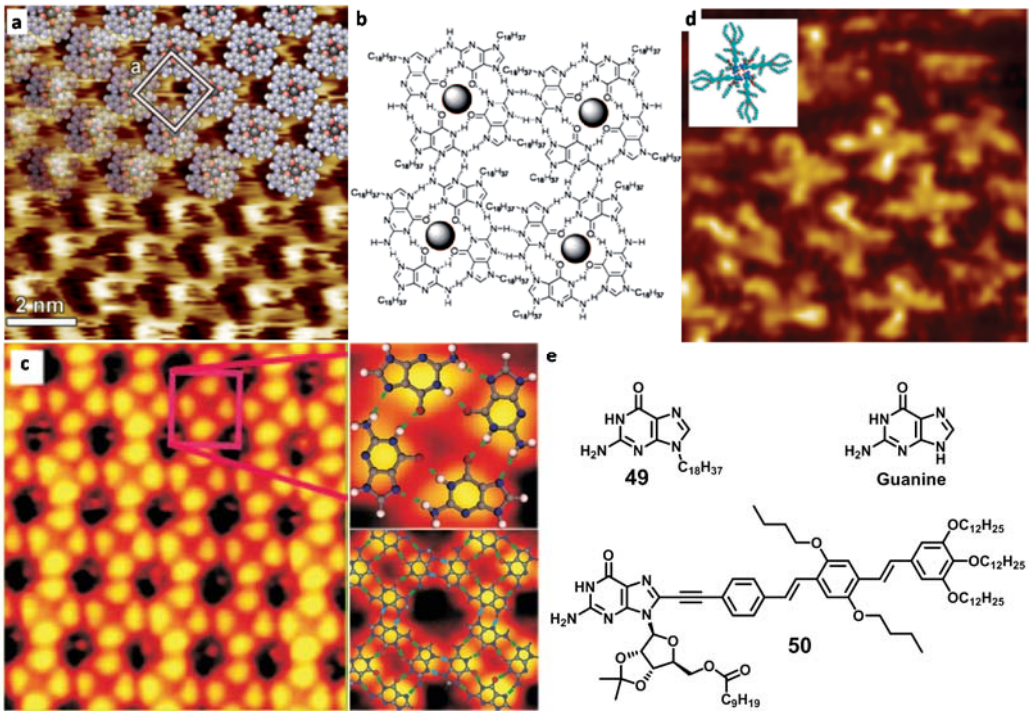


Figure 29. STM images of different organizations undergone by G derivatives. (**a**) STM image of **49** at the TCB/HOPG interface (I_t = 5 pA, V_s = 200 mV) in the presence of K^+, along with (**b**) association mode. (**c**) UHV-STM image of guanine on Au (111) (8 x 8 nm^2, 150–170 K) along with magnifications. (**d**) STM image of **50** at the 1-phenyloctane/HOPG interface (I_t = 28 pA, V_s = -0.600 mV) along with model. (**e**) Chemical structures of the compounds.

Different G analogs present interesting aggregation phenomena as well. Among them *iso*guanine (iG) or 2-hydroxyadenine, which is an isomer of guanine and product of oxidative damage to DNA, displays fascinating supramolecular chemistry. Davis and colleagues studied in depth the self-assembly of *iso*guanosine derivatives.[128] They were able to self-assemble compound **51** into cyclic structures, whose size could be tuned by altering the templating metal cation and/or the extraction conditions. Low concentrations of Na^+ afforded *iso*guanosine tetramers (Figure 30b) in organic solvents, more stable than the famous G_4.[128a] The four oxygens in the central cavity supplied this species with high affinity for cations. In the presence of K^+ or Ba^{2+}, **51** formed very robust octamers,[128b] proven to be remarkable selective ionophores for Cs^+.[129] On the contrary, compound **51** gave a pentameric assembly in the presence of Cs^+ (Figure 30c), where the nucleobases

[128] a) J. T. Davis, S. Tirumala, J. R. Jenssen, E. Radler, D. Fabris, *J. Org. Chem.* **1995**, *60*, 4167–4176; b) S. Tirumala, J. T. Davis, *J. Am. Chem. Soc.* **1997**, *119*, 2769–2776; c) A. L. Marlow, J. T. Davis, *Tetrahedron Lett.* **1999**, *40*, 3539–3542.
[129] J. T. Davis, S. Tirumala, A. L. Marlow, *J. Am. Chem. Soc.* **1997**, *119*, 5271–5272.

were bound through their Watson-Crick and Hoogsteen faces, forming an angle of 72º. Additionally, when a liquid-liquid extraction was performed in the presence of NaI or CsI, a sandwich complex corresponding to a decamer was observed (Figure 30d).[128c]

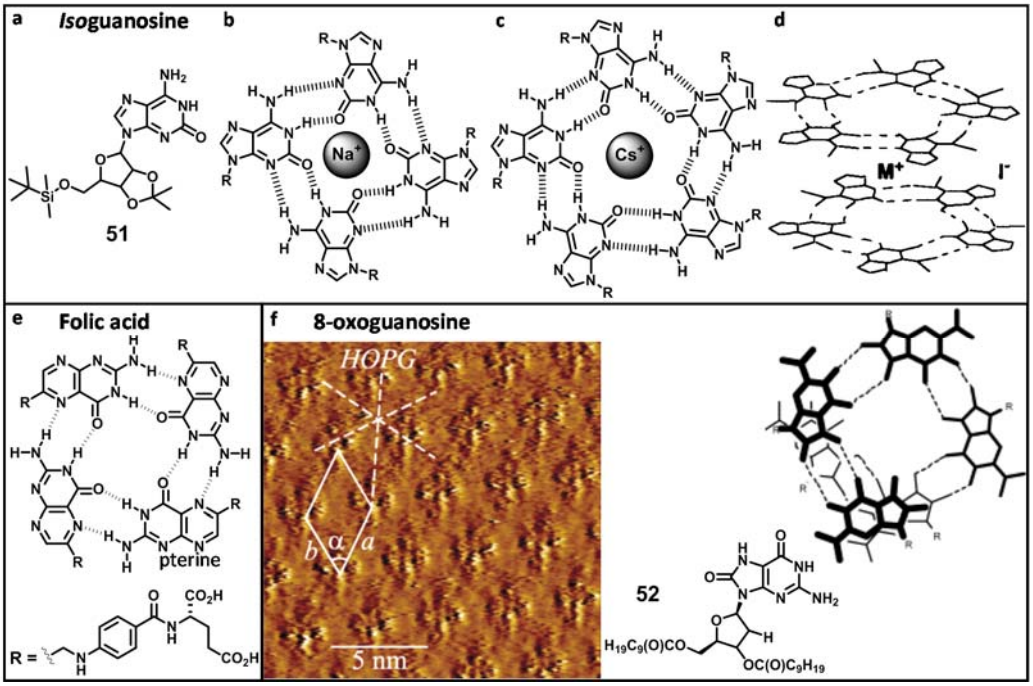


Figure 30. Supramolecular chemistry undergone by different G analogues. (**a**-**d**) Tetramer, pentamer and decamer formed by **51**, along with chemical structure. (**e**) Tetramer formed by folic acid and chemical structure. (**f**) STM image of the 8-oxoguanosine **52** at the TCB/HOPG interface (21 pA, 0.684 V) along with chemical structure and model of the helical structure.

Another example is folic acid, which is a B vitamin and an important biologically active molecule. This G analog carries a pterine heterocycle and the group of Spada showed how it self-associated into tetrameric macrocycles, stabilized through Hoogsteen-type H-bonding and stacking (Figure 30e).[130] Gottarelli and coworkers also studied the case of 8-oxoguanosine **52**, which self-organized into supramolecular helical architectures in the liquid crystalline phase, in solution and onto surfaces (Figure 30f).[131]

Mixtures of nucleobases have been less investigated on the surface, as the different adsorption of the components on the substrate may suppose an additional difficulty for imaging. The different pattern followed by each of the Watson-Crick pairs should allow for

[130] a) S. Bonazzi, M. M. DeMorais, G. Gottarelli, P. Mariani, G. P. Spada, *Angew. Chem. Int. Ed.* **1993**, *32*, 248–250; b) F. Ciuchi, G. Di Nicola, H. Franz, G. Gottarelli, P. Mariani, M. G. Ponzi Bossi, G. P. Spada, *J. Am. Chem. Soc.* **1994**, *116*, 7064–7071; c) G. Gottarelli, E. Mezzina, G. P. Spada, F. Carsughi, G. Di Nicola, P. Mariani, A. Sabatucci, S. Bonazzi, *Helv. Chim. Acta* **1996**, *79*, 220–234.

[131] T. Giorgi, S. Lena, P. Mariani, M. A. Cremonini, S. Masiero, S. Pieraccini, J. P. Rabe, P. Samori, G. P. Spada, G. Gottarelli, *J. Am. Chem. Soc.* **2003**, *125*, 14741–14749.

some selectivity in the interactions. Indeed, the G-nucleobase shows strong affinity for binding in solution with the C-base, but not with U. Also, as explained in section **1.1.**, the Watson-Crick *DAA–ADD* G–C base-pair is much stronger than the *DAD–ADA* 2-aminoadenine–U base-pair, because of the secondary electrostatic interactions. The complementarity between nucleobases appears as a potential tool for the design of finely-tuned self-assembling nanostructures. However, when concentrated on the surface, non-complementary unions can occur, as multiple H-bonds are available. Besenbacher and co-workers demonstrated the stability of the Watson-Crick G–C base-pair (Figure 31a-c),[132] but also the viability of the weaker non-complementary A–C and G–U pairs, under UHV conditions[132a] and at the solid-liquid interface,[133] respectively. After sequential co-deposition of C, and then G, onto Au (111), an increase in the number of bicomponent fivefold rings was observed; while nothing of the sort occurred when depositing adenine (even if H-bonded binary arrays of A and C were still observed).[132a] When heating the complementary C + G mixture, the network remained disordered, whereas the non-complementary C + A mix segregated into A islands and C zigzag branches. Upon mixing G and U molecules, a self-assembled nanoscale patterned supramolecular structure, consisting of a parallel arrangement of 1D chains of G–U cyclic units, was observed (Figure 31d). The structure was stabilized by strong H-bonding along the chains and relatively weaker interactions between them. The authors could investigate the formation mechanism and proposed that first, a strong G–U dimer was created and then, two dimers formed a cyclic structure (Figure 31f), which acted as the fundamental building block of the monolayer. Later, the same research group went further in the stabilization on the surface of the Watson-Crick G–C pair and prepared the corresponding *N*-arylated derivatives (Figure 31h).[132b] They could observe how the modified nucleobases self-assembled into a porous nanostructure on Au (111), through complementary H-bonding (Figure 31g).

[132] a) R. Otero, W. Xu, M. Lukas, R. E. A. Kelly, E. Laegsgaard, I. Stensgaard, J. Kjems, L. N. Kantorovich, F. Besenbacher, *Angew. Chem. Int. Ed.* **2008**, *47*, 9673–9676; b) W. Xu, J.-g. Wang, M. F. Jacobsen, M. Mura, M. Yu, R. E. A. Kelly, Q.-q. Meng, E. Laegsgaard, I. Stensgaard, T. R. Linderoth, J. Kjems, L. N. Kantorovich, K. V. Gothelf, F. Besenbacher, *Angew. Chem. Int. Ed.* **2010**, *49*, 9373–9377.

[133] W. Mamdouh, R. E. A. Kelly, M. Dong, L. N. Kantorovich, F. Besenbacher, *J. Am. Chem. Soc.* **2008**, *130*, 695–702.

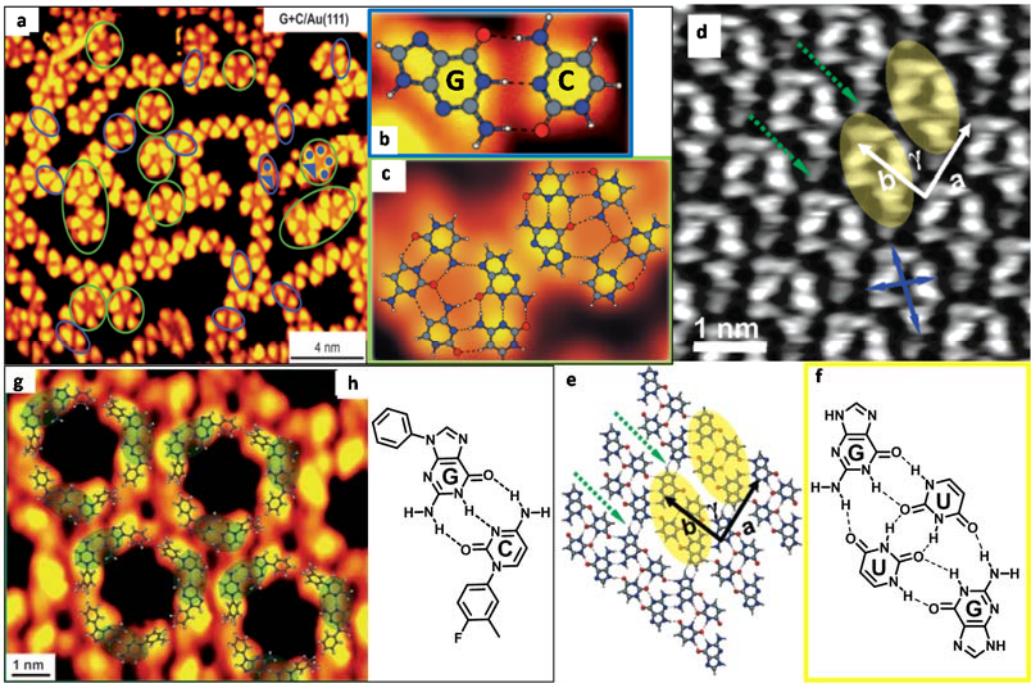


Figure 31. STM images of the self-assembly of different G–nucleobase pairs. (**a**) UHV-STM image of a G + C mixture on Au (111) along with (**b**,**c**) magnifications. (**d**) High-resolution STM image of G–U-base pairs at the 1-octanol/graphite interface (I_t = 0.70 nA, V_{bias} = 589.0 mV). (**e**) Molecular structure proposed by *ab initio* calculations. Yellow ovals indicate the unit cell, whose size is indicated by blue arrows and the H-bonds between the G–U-cycles by green arrows. (**f**) Unit cell chemical structure. (**g**) DFT-calculated molecular model superimposed on the close-up high-resolution STM image of a self-assembled porous network formed by sequential co-deposition of *N*-aryl-modified nucleobases C and G. (I_t = -1.7 nA, V_t = -2500 mV). The Watson-Crick G–C pairs are highlighted by green ellipsoids and detailed in (**h**).

2.4. Chirality on the surface.

The formation of chiral domains on solid substrates will be discussed in Chapters 2 and 3 of this Thesis. Therefore, we decided to include a short section to introduce the topic "*Chirality on the Surface*". The importance of controlling and understanding the self-assembly of molecular tectons into non-superposable arrays in two dimensions will be here addressed.

Chirality is present at various levels from molecular to supramolecular, macroscopic and even on the galactic scale. Controlling chirality is crucial for the chemical industry and for many chemical processes, such as catalysis or crystallization, and self-assembly tries to manage it. Understanding its behavior at the nanoscale would be an enormous step toward reaching control at the macroscale. Therefore, much effort has been put in this area for several decades now and different approaches keep being investigated. Crystallization constitutes the most economical procedure to obtain enantiopure compounds industrially.

In fact, a racemate generally condensates in three ways: (1) as a *racemic compound*, in which the two enantiomers are present in the same condensate; (2) as a *conglomerate*, in which molecules form condensates of only one enantiomer, but where the sample stays racemic, since it contains equal amounts of enantiomorphic condensates; and (3) as a *pseudoracemate*, in which the condensates contain both enantiomers in a non-ordered arrangement.[134] Among these three forms, only the conglomerate implies spontaneous resolution, which is the segregation of enantiomers upon crystallization.[135] Although convenient, the separation of the 3D-crystals composing a conglomerate is not always possible, because its formation is, actually, not very frequent and even less predictable.[136] Therefore, new alternative ways to obtain enantiomeric crystals must be explored and, once again, nature can serve as source of inspiration.

Chiral supramolecular architectures, which arise from the non-symmetric spatial arrangement of the components, are a common feature in biological systems. Supramolecular chirality is strongly related to both the chirality of the molecular building blocks and their assembly manner. Actually, all kinds of molecules (chiral or achiral) may be possibly assembled into chiral nano- or microstructures through non-covalent bonds. In fact, the induction of chirality in achiral condensed phases is also of great importance in artificial systems like, for example, liquid crystal mesophases. The challenge to understand how chirality, at the supramolecular level, is related to the molecular components remains intriguing.[137] In this context, surface chirality stands out, having an intimate relationship with the origin of chirality in nature, chiral molecular recognition or enantioselective catalysis.[138] By bringing 3D crystals down to the interface, the investigation of the behavior of chirality is simplified. This study of the expression of chirality in 2D molecular crystals (at the level of a self-organized molecular monolayer) can provide a good understanding of how chiral information can be transmitted from individual molecular building blocks, to the second and eventually to the third dimension.[139]

A structure is chiral in 2D if it is non-superposable to its mirror image by rotation or by translation in the plane of the surface. One possibility is to consider chiral building blocks. In this scenario, chiral molecules form, most frequently, chiral supramolecular structures on achiral surfaces. For example, Guo et *al.*[140] showed how the achiral OPV derivative (**53**) formed an ordered racemic conglomerate monolayer composed of clockwise (CW) and

[134] L. Pérez-García, D. B. Amabilino, *Chem. Soc. Rev.* **2002**, *31*, 342–356.
[135] L. Pasteur, *C. R. Seances Acad. Sci.* **1848**, *26*, 535–539.
[136] a) J. Jacques, A. Collet, S. H. Wilen, *Enantiomers, Racemates and Resolutions*, Krieger Publishing Company, Malabar, Florida, **1994**; b) C. Pratt-Brock, J. D. Dunitz, *Chem. Mater.* **1994**, *6*, 1118–1127; c) R. G. Kostyanovsky, A. P. Avdeenko, S. A. Konovalova, G. K. Kadorkina, A. V. Prosyanik, *Mendeleev Commun.* **2000**, *10*, 16–18.
[137] L. Zhang, L. Qin, X. Wang, H. Cao, M. Liu, *Adv. Mater.* **2014**, *26*, 6959–6964.
[138] a) S. M. Barlow, R. Raval, *Surf. Sci. Rep.* **2003**, *50*, 201–342; b) M. O. Lorenzo, C. J. Baddeley, C. Muryn, R. Raval, *Nature* **2000**, *404*, 376–379; c) G. A. Attard, *J. Phys. Chem. B* **2001**, *105*, 3158–3167; d) M. M. Knudsen, N. Kalashnyk, F. Masini, J. R. Cramer, E. Laegsgaard, F. Besenbacher, T. R. Linderoth, K. V. Gothelf, *J. Am. Chem. Soc.* **2011**, *133*, 4896–4905.
[139] N. Katsonis, E. Lacaze, B. L. Feringa, *J. Mater. Chem.* **2008**, *18*, 2065–2073.
[140] Z. Guo, I. De Cat, B. Van Averbeke, E. Ghijsens, J. Lin, H. Xu, G. Wang, F. J. M. Hoeben, Z. Tomovic, R. Lazzaroni, D. Beljonne, E. W. Meijer, A. P. H. J. Schenning, S. De Feyter, *J. Am. Chem. Soc.* **2013**, *135*, 9811–9819.

counter-clockwise (CCW) rosette domains at the 1-phenyloctane/HOPG interface, whereas the (*S*)-derivative **46** (see Figure 27c,d, section **2.2.** in this Introduction) only formed CW rosettes and the (*R*)-enantiomer **54** only CCW rosette-assemblies. However, the stereochemistry of a self-assembled monolayer can be imposed through the competition between molecule-molecule and molecule-substrate interactions. De Feyter *et al.*[141] demonstrated, for instance, how the expression of chirality in monolayers formed by enantiomerically pure isophthalic acid molecules in solution was not a general rule and depended not only on the nature of the solvent and substrate, but also on the position of the stereogenic centers with respect to the plane of the substrate.

Another possibility comes from achiral molecules that become chiral by addition of substituents to one or both their heterotopic faces (*i.e.* by desymmetrization with respect to the main plane of the molecules): these are said prochiral. Consequently, many planar molecules are prochiral: they do not possess a symmetry axis, but when adsorbed on an achiral surface lead to the formation of localized 2D chiral structures. The spontaneous formation of enantiopure or racemic surface supramolecular patterns from the adsorption and assembly of chiral or prochiral molecules has also been well-documented.[142] Again, Guo *et al.*[143] studied in 2011 by STM the self-assembly of the OPV derivative **53** at the 1-octanol/HOPG interface and the effect of the addition *D*- and *L*-thymidine. Upon addition of the pyrimidine, a morphological change of the **53** monolayer pattern from CW and CCW rosettes (Figure 32a) to H-bonded thymidine-**53** dimers was observed. Moreover, an overall chiral monolayer was induced by the introduction of thymidine enantiomers, where the chirality of the dimers and the monolayer was dominated by the chirality of nucleobase and enantiomorphous dimers were obtained from **53** with *D*-/*L*-thymidine (Figure 32b,c). Thymidine molecules were proven to play a key role in not only the chiral expression but also the surface packing. Later, Chen *et al.*[144] investigated the co-assembly onto HOPG of 5-(benzyloxy)isophthalic acid derivatives (**55**-**C**$_n$) with different chiral and achiral co-adsorber alkyl alcohols. When co-adsorbed with 1-heptanol and 1-octanol achiral molecules, **55** formed both CW and CCW domains. However, when mixed with chiral alcohols, single chirality domains were obtained (Figure 32e). The chirality of the supramolecular assembly followed an odd–even alternation rule with respect to the position of the stereogenic center owing to the alternation of the spatial configuration of the stereocenter when the co-adsorber is physisorbed on a surface.

[141] S. De Feyter, P. C. M. Grim, M. Rucker, P. Vanoppen, C. Meiners, M. Sieffert, S. Valiyaveettil, K. Müllen, F. C. De Schryver, *Angew. Chem. Int. Ed.* **1998**, *37*, 1223–1226.
[142] a) R. Fasel, M. Parschau, K. H. Ernst, *Nature* **2006**, *439*, 449–452; b) Q. Chen, N. V. Richardson, *Nat. Mater.* **2003**, *2*, 324–328; c) M. Bohringer, K. Morgenstern, W. D. Schneider, R. Berndt, *Angew. Chem. Int. Ed.* **1999**, *38*, 821–823; d) S. Weigelt, C. Busse, L. Petersen, E. Rauls, B. Hammer, K. V. Gothelf, F. Besenbacher, T. R. Linderoth, *Nat. Mater.* **2006**, *5*, 112–117.
[143] Z. Guo, I. De Cat, B. Van Averbeke, J. Lin, G. Wang, H. Xu, R. Lazzaroni, D. Beljonne, E. W. Meijer, A. P. H. J. Schenning, S. De Feyter, *J. Am. Chem. Soc.* **2011**, *133*, 17764–17771.
[144] T. Chen, S.-Y. Li, D. Wang, M. Yao, L.-J. Wan, *Angew. Chem. Int. Ed.* **2015**, *54*, 4309–4314.

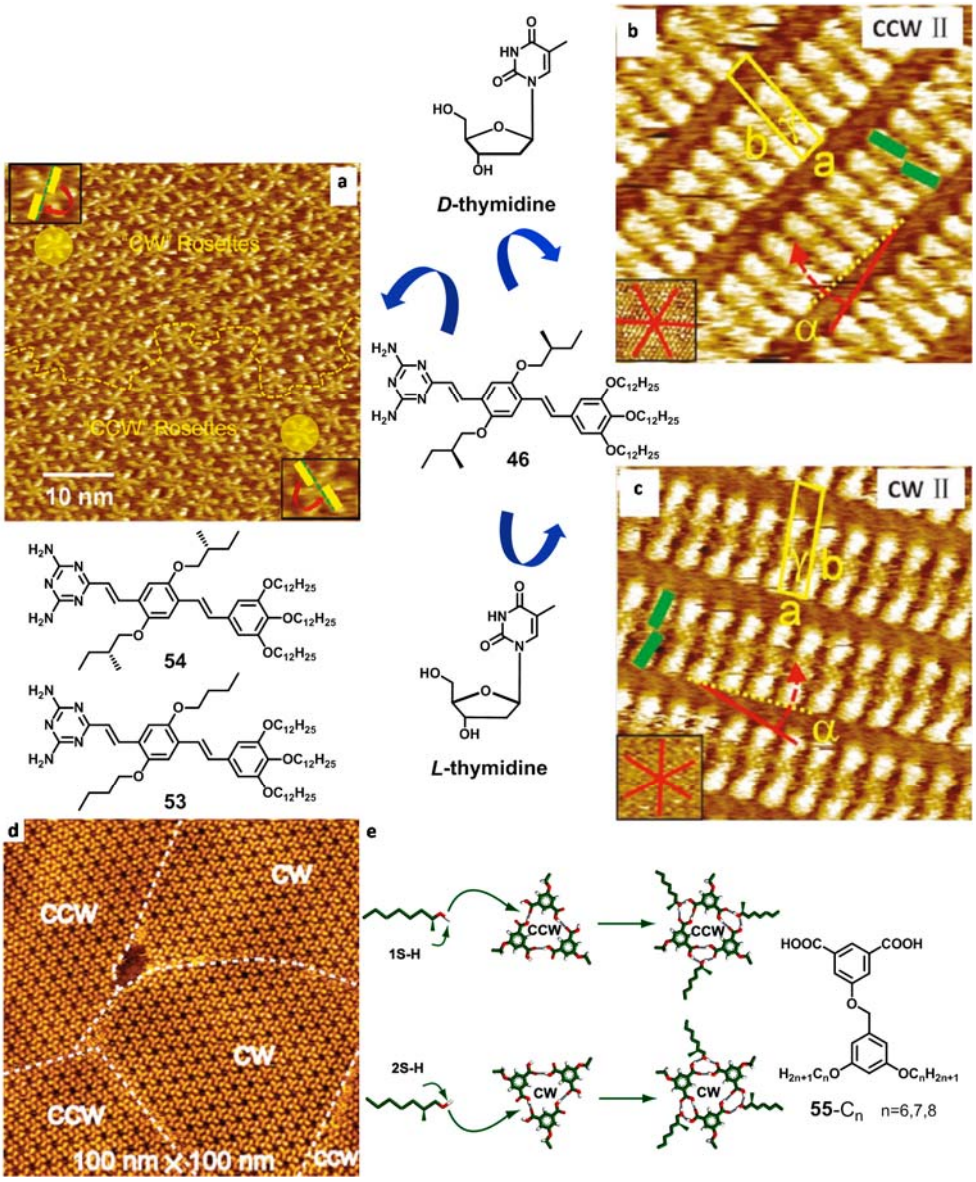


Figure 32. Investigations on the expression of chirality on the surface. (**a**) STM image of **46** at the 1-octanol/HOPG interface (I_{set} = 0.28 nA, V_{bias} = -0.30 V) showing the structure of CW and CCW rosettes. (**b**) STM image of **46** with *D*-thymidine (I_{set} = 1.0 nA, V_{bias} = -0.25 V) showing the structure of CCW dimers. (**c**) **46** with *L*-thymidine (I_{set} = 0.189 nA, V_{bias} = -0.302 V) showing CW dimers. (**d**) Typical STM image of the enantiomorphous networks of **55**-C_6 co-assembled with 1-heptanol. (**e**) Illustration of the remote chiral transmission in the co-assembly of **55**-C_n derivatives with alcohols 1*S*-H and 2*S*-H through H-bonding, along with the general chemical structure of **55**.

Background and Objectives.

Background in *MSMn.*

1) Nanostructured Molecular Systems and Materials.

This Thesis has been carried out in the *Nanostructured Molecular Systems and Materials* (*MSMn*) group at the Universidad Autónoma de Madrid. The research in *MSMn* aims at a common final objective: to improve or create new functions in organic materials by rationally ordering functional molecules at the nanoscale using the tools of self-assembly, or in other words, using noncovalent synthesis. One of the main projects in the group, *Programmed Nanostructuration of Organic Materials*, employs the structure of a self-assembled cyclic tetramer as a common feature in different subprojects (Figure 33): (1) the study of non-covalent macrocyclizations; (2) the development of π-conjugated functional self-assembled systems; (3) the formation of nanostructured porous surfaces and networks; (4) the polymerization of these cyclic tetramers into nanotubes with custom-tailored pores; and (5) the transfer of this supramolecular motif to biopolymers, to create tubular quadruplex DNA.

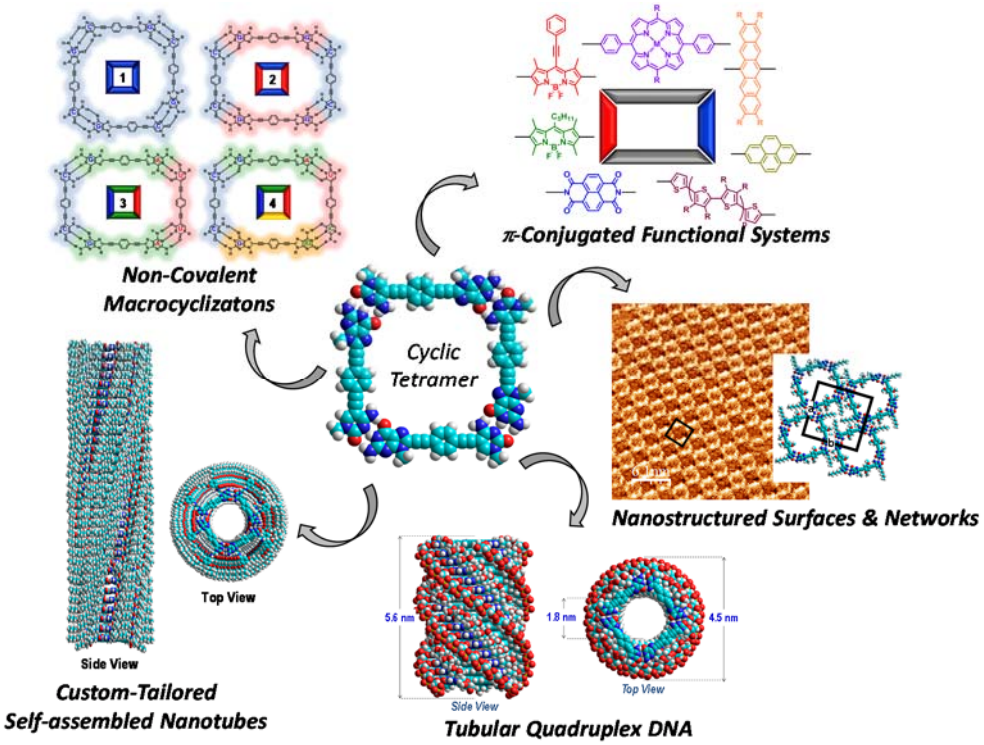


Figure 33. Programmed Nanostructuration of Organic Materials.

2) The Cyclic Tetramer.

This common macrocycle is built *via* H-bonding interactions between four monomer molecules (Figure 34). The monomers basically comprise a *rigid π-conjugated* ***central block*** that is linearly disubstituted at both ends through suitable ***spacers*** with *self-assembling* ***directors,*** able to interact by H-bonding.

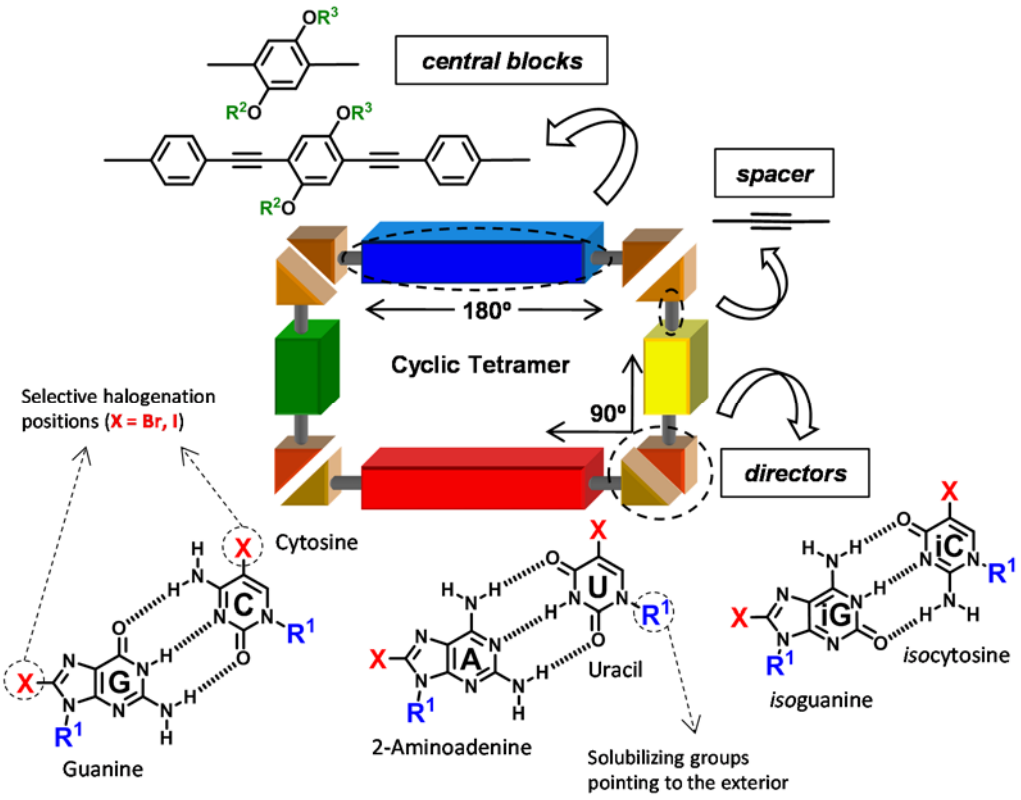


Figure 34. Key structural elements that lead to cyclic tetramer assembly.

The H-bonding ***directors*** are actually nucleobases, *i.e.* naturally occurring guanine (G), cytosine (C) and uracil (U); and non-natural 2-aminoadenine (A),[145] *iso*guanine (iG) and *iso*cytosine (iC). Together, they constitute a family of three complementary Watson-Crick couples: G–C, A–U, and iG–iC, that associate *via* triple H-bonding patterns. That is the reason why adenine has been replaced by 2-aminoadenine, so as to keep the three-fold H-bonding pattern and stabilize self-assembled structures. When these bases are subjected to electrophilic halogenation reactions, purines are substituted at the 8-position and pyrimidines at the 5-position (marked as **X** in Figure 34). It is essential to note that when the purine-pyrimidine pairs interact *via* complementary Watson-Crick H-bonding, these

[145] 2-aminoadenine will be abbreviated as A for the sake of simplicity.

two positions form an exact 90º angle. On the other hand, the marked **R[1]** positions, the 9-position in purines and the 1-position in pyrimidines, are always pointing toward the exterior of the nucleobase pairs. These are the same positions where the riboses are placed in RNA, and they can be further utilized to incorporate key solubilizing groups in order to approach different purposes for the cyclic tetramers.

The ***central blocks*** have to be rigid units substituted at both ends with an exact angle of 180º. They will typically comprise π-conjugated units that carry reactive groups at opposite positions and that can be endowed with diverse functions and equipped with different lateral groups (**R[2]** and **R[3]** in Figure 34), depending on the application requirements. There is a large number of functional π-conjugated units that fulfill the prerequisite of linear di-substitution: TTF, conjugated oligomers, aromatic acenes, perylenes, ABAB porphyrin or phthalocyanine macrocycles, etc (Figure 35).

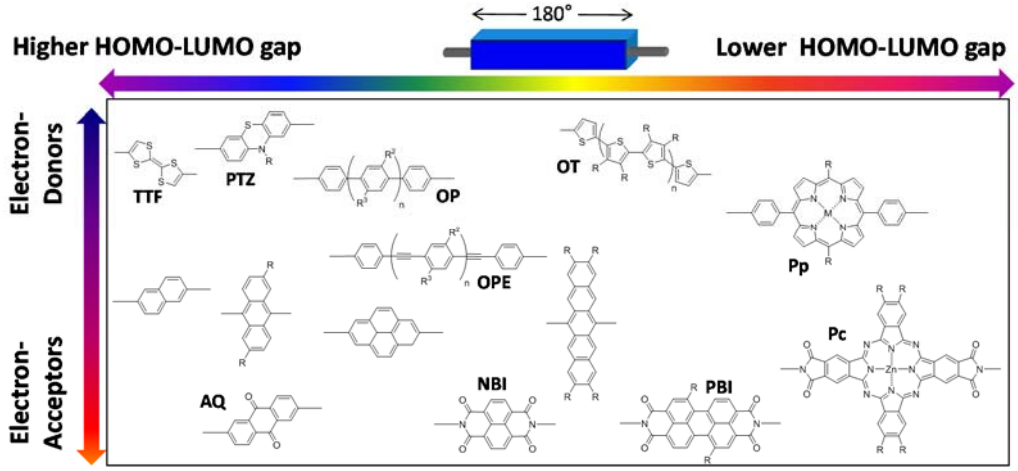


Figure 35. Potential π-conjugated functional blocks.

Ethynyl groups have been selected as the ***spacers*** between the different units forming the monomer. Such groups are interesting to us because they are linear, rigid and allow for substitution in this suitable 180º angle. They have some rotational liberty and a minimum steric hindrance, which is necessary for a sufficient conformational freedom. They are also π-conjugated and can facilitate the electronic coupling between the different units in the cyclic tetramers. Finally, they possess a wide chemical versatility, mainly in simple and mild cross-coupling reactions,[146] and allow for a convenient coupling between central block and nucleobases through Sonogashira reactions, which are central in this project. The nucleobases could also be directly connected without intercalating any central block, through an ethynyl spacer or just a single covalent bond.

[146] R. Chinchilla, C. Nájera, *Chem. Rev.* **2014**, *114*, 1783–1826.

Overall, the linear structure of the dinucleobase monomer together with the 90º angle imposed by H-bonding interactions at the edges will lead to rectangular assemblies composed of four molecules.

3) Supramolecular Macrocyclization in Solution.

The assembly of dinucleosides into H-bonded macrocycles has been extensively investigated in our research group. A first simple ditopic monomer comprising a *p*-diethynylbenzene ring disubstituted with complementary G and C nucleobases at the edges was prepared (**GC$_{rib1}$**; Figure 36). Bulky, lipophilic ribose groups were placed at the **R^1** nucleobase positions to afford solubility and prevent stacking interactions, in order to focus on the cyclotetramerization process by H-bonding, which was primarily studied by NMR and optical spectroscopic techniques.

In the self-assembly process undergone by these monomers, the sticky ends recognize each other and bind forming a dimer with an association constant that corresponds to the G–C Watson-Crick pairing (K_{ref}). Then, further association will lead to a trimer, a tetramer and so on: this process will continue obtaining a distribution of supramolecular oligomers as long as the thermodynamics of the system allows it (Figure 36). An interplay between enthalpy (ΔH) and entropy (ΔS) takes place, where the association is enthalpically favored (G–C binding is exothermic) but entropically disfavored (ΔS decreases upon molecular association). These magnitudes can be influenced as a function of the temperature, the concentration and the solvent environment. In our system, Watson-Crick pairing provides a 90º angle and our monomers have been designed to be very rigid (they can only rotate around the sigma bonds and alternate between *anti* and *syn* conformations) (Figure 36). Upon association, open supramolecular oligomers can be formed by participation of "*anti*" conformations, where the Watson-Crick edges are in opposite sides. In contrast, the exclusive participation of *syn* conformations leads to cyclic species, the cyclic tetramer being the most stable one in solution, since it is the smallest, non-strained macrocycle. In other words, the monomer is preorganized to form a cyclic tetramer. The stabilization gained when going from an open to a cyclic system is represented by the product K_{ref} x *EM*, where K_{ref} is the reference G–C association constant (since there is an additional binding site) and *EM* stands for the *Effective Molarity* (see Introduction) and takes into account that the last binding event to form the cycle is intramolecular (or, formally speaking, intraspecies). This is known as *Chelate Cooperativity* and the *EM* is defined by the ratio between intra- (K_{intra}) and intermolecular (K_{inter}) equilibrium constants and represents how favorable is the intramolecular process with respect to the intermolecular one. The monomers have been carefully designed for an efficient supramolecular macrocyclization in solution, hence leading to discrete cyclic tetramer assemblies (Figure 36). However, depending on the experimental conditions, these macrocycles could be in equilibrium with the corresponding monomers and open oligomers, formed by rotation of the block-base single bonds.

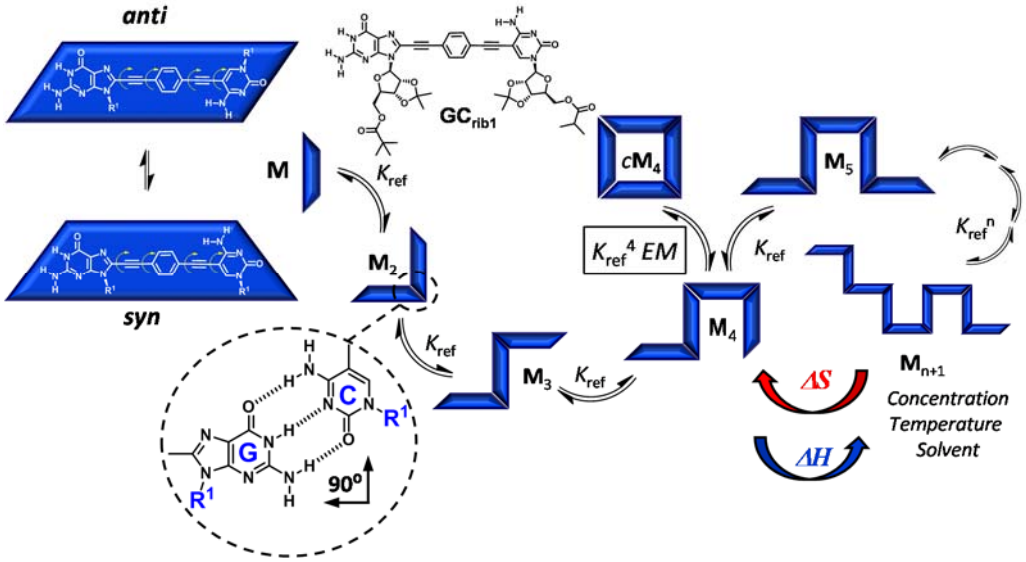


Figure 36. The cyclotetramerization equilibrium process.

The experimental results obtained for **GC$_{rib1}$** indicate that the cyclic tetramers are formed quantitatively in apolar solvents, as determined by 1D and 2D NMR experiments, DOSY, and ESI Q-TOF MS.[147] Three different approaches were then employed to dissociate them and study the underlying equilibrium processes: (1) addition of highly polar solvents, like DMSO or DMF, (2) dilution experiments, and (3) competition experiments with a mononucleoside (**S**).[148] The conclusions drawn from these combined experiments revealed that **GC$_{rib1}$** cyclic tetramer exhibits an impressive thermodynamic stability and constitute a kinetically steady product in the overall self-assembly landscape, even in highly polar solvents like DMF where H-bonded association is typically too weak. The titration experiments with a mononucleoside competitor, were particularly revealing (Figure 37). Upon addition of the cytidine stopper **S** to a solution of (**GC$_{rib1}$**)$_4$ a gradual disappearance of the tetramer proton signals and the emergence of a new set of signals attributed to the **GC$_{rib1}$-S** complex, in fast equilibrium with excess **S** are observed (Figure 37b). This process could be equally followed by fluorescence titration experiments (Figure 37c) monitoring the emission of molecule **GC$_{rib1}$** in the two associated states. Figure 37d plots the fraction of **GC$_{rib1}$** associated as a cyclic tetramer *vs* the number of equivalents of the **S** stopper added in three solvent systems: DMF, THF and $CHCl_3$. In order to fully dissociate the macrocycle, *ca.* 50 eq of **S** must be added, which underlines the stability of the macrocyclic assembly.

[147] a) C. Montoro-García, J. Camacho-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, M. J. Mayoral, D. González-Rodríguez, *Angew. Chem. Int. Ed.* **2015**, *54*, 6780–6784 (VIP Paper); b) S. Romero-Pérez, J. Camacho-García, C. Montoro-García, A. M. López-Pérez, A. Sanz, M. J. Mayoral, D. González-Rodríguez, *Org. Lett.* **2015**, *17*, 2664−2667.
[148] J. Camacho-García, C. Montoro-García, A. M. López-Pérez, N. Bilbao, S. Romero-Pérez, D. González-Rodríguez, *Org. Biomol. Chem.* **2015**, *13*, 4506–4513.

The cyclotetramerization constants and *EM* values were determined in the different solvent systems by diverse spectroscopic methods. The competition trends and the calculated *EM* values do not seem to reveal a clear relationship between *EM* and binding strength (K_{ref}), modulated by the solvent. The rigidity and non-rotatable nature of the multipoint G–C binding interaction appeared as a key factor that notably increases the magnitude of *EM* in $(\mathbf{GC_{rib1}})_4$ (10^2–10^3 M) when compared to other cyclic tetramers based on metal–ligand interactions (*EM* = 0.1–20 M).

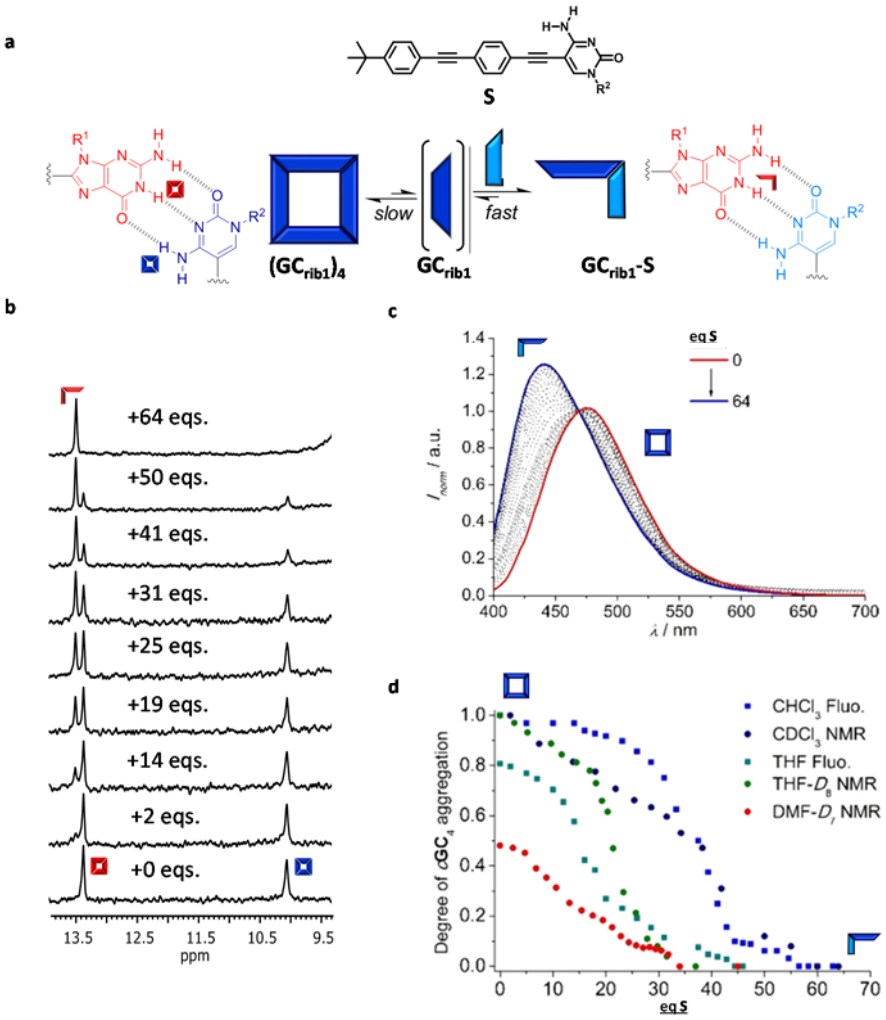


Figure 37. **Evaluation of the stability on the cyclic tetrameric species in solution by competition experiments.** (**a**) Titration experiments of $(\mathbf{GC_{rib1}})_4$ with **S** at 298 K measured by (**b**) ^{1}H NMR ($C = 10^{-3}$ M) in $CDCl_3$ or (**c**) emission spectroscopy ($C = 5 \times 10^{-5}$ M; $\lambda_{exc} = 390$ nm) in $CHCl_3$. (**d**) Plots of the degree of $(\mathbf{GC_{rib1}})_4$ association, measured by ^{1}H NMR or emission, as a function of the equivalents of **S** added in DMF, THF or $CHCl_3$.

Aside from $(\mathbf{GC_{rib1}})_4$, the quantitative formation of remarkably stable cyclic tetramers $(\mathbf{iGiC_{rib1}})_4$ and $(\mathbf{AU_{rib1}})_4$ in solution has also been demonstrated and characterized by different experiments (Figure 38).[149]

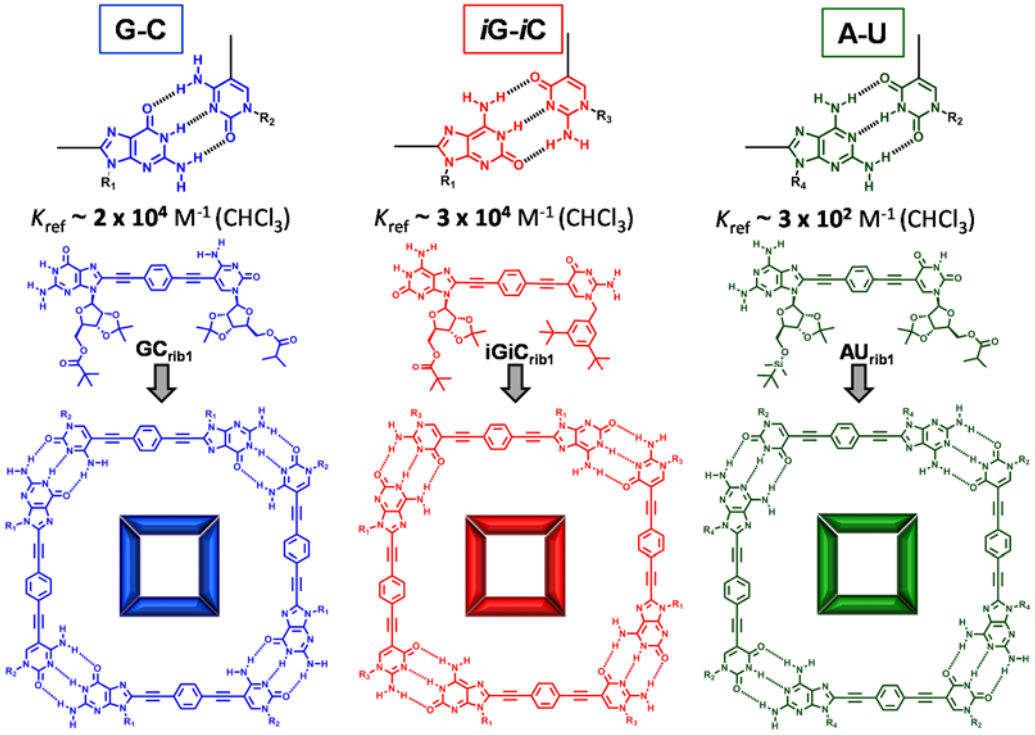


Figure 38. Selective cyclic tetramer assembly through nucleobase H-bonding self-recognition.

The self-assembly of **iGiC$_{rib1}$** and **AU$_{rib1}$** was analyzed in different solvents by a wide number of concentration- and temperature-dependent spectroscopic methods (1D and 2D ^{1}H NMR, as well as absorption, emission and CD spectroscopy). The equilibrium constants of the cyclotetramerization processes (K_T) could be obtained, together with the reference association constants between complementary mononucleosides (K_{ref}).[149]

The set of results demonstrated that **AU$_{rib1}$** forms considerably less stable cyclic tetramers than **GC$_{rib1}$** or **iGiC$_{rib1}$**, which, on the other hand, revealed comparable qualitative and quantitative association behavior. This stability trend was expected, since the individual A–U binding constant (K_{ref} ~ 3 x 10^2 M^{-1}) is typically about 2 orders of magnitude lower than G–C or iG-iC (K_{ref} ~ 3 x 10^4 M^{-1}) in $CHCl_3$.[148] This can be explained by the Jorgensen model[22a] (see Introduction) and is due to the different stabilizing/destabilizing secondary H-bonding interactions between vicinal donor and acceptor groups in the *DAD–ADA* (A–U) pair *vs* the *DDA–ADD* (G–C or iG–iC) pair. However, our experimental results

[149] C. Montoro-García, J. Camacho-García, A. M. López-Pérez, M. J. Mayoral, N. Bilbao, D. González-Rodríguez, *Angew. Chem. Int. Ed.* **2016**, *55*, 223–227.

suggested that such decrease in stability was actually much larger. In fact, the K_T for **AU$_{rib1}$** is not only reduced by a decrease in K_{ref}, but also by a substantial decrease in the magnitude of *EM*. Actually, the symmetry of the *DAD–ADA* bonding pattern introduces the possibility of multiple binding modes (*i.e.* Watson-Crick and reverse Watson-Crick H-bonding) and hence a higher number of degrees of freedom in linear oligomers, which are then lost upon macrocyclization (Figure 39). This effect, of entropic origin, has a large impact on the *EM* of the system, which in our case is reduced by about 3 orders of magnitude.

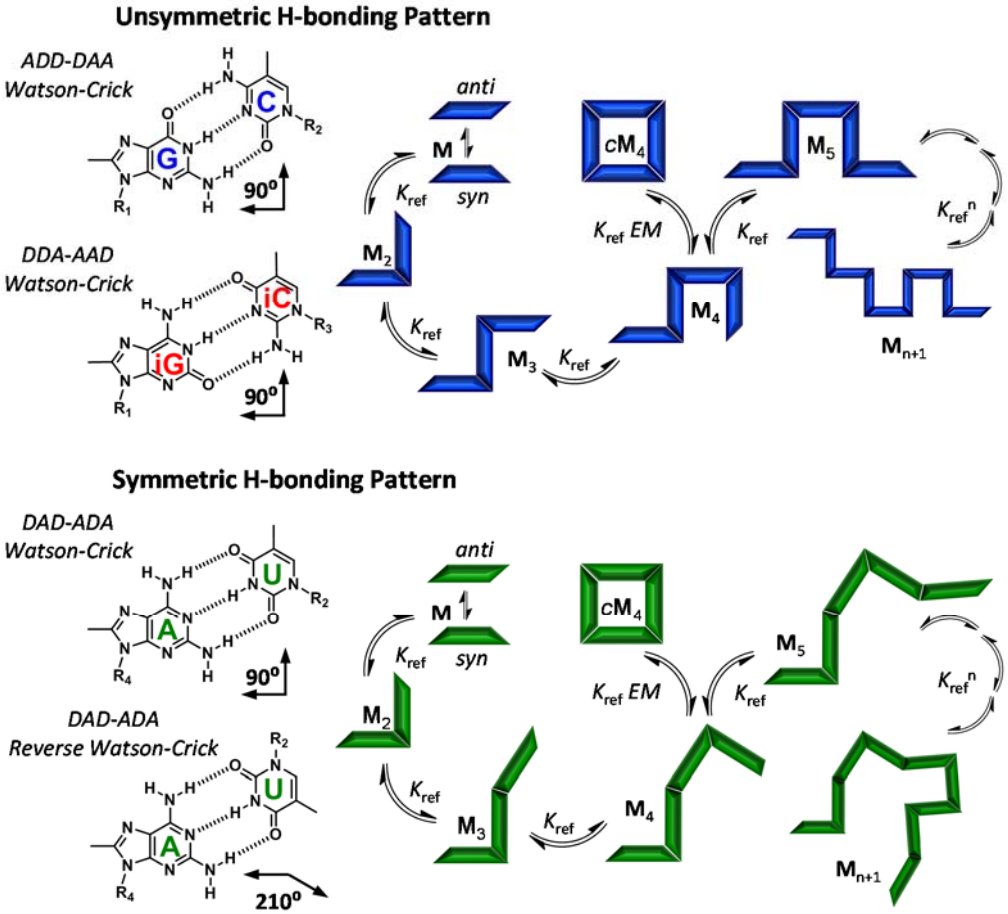


Figure 39. **Unsymmetric *vs* Symmetric H-bonding patterns.** *ADD–DAA* (G–C) and *DDA–ADD* (iG–iC) H-bonding patterns are unsymmetric and thus formed with a 90º angle. In contrast, the *DAD–ADA* symmetric pattern (A–U) can be formed with 90º or 210º angles, hence leading to supramolecular polymers with a wider number of conformational possibilities that are lost upon cyclotetramerization.

4) Self-sorting in Solution.

Self-sorting in supramolecular chemistry is defined as high fidelity recognition between molecules (and ions) within complex mixtures.[150] If affinity for others is shown, this assembly process is called self-discrimination (social self-sorting),[151] and the affinity for itself would be called self-recognition (narcissistic self-sorting)[152] (Figure 40).[153] Self-sorting systems can, in turn, be subdivided into those displaying thermodynamic or kinetic self-sorting, depending if they have reached a thermodynamic equilibrium or can be considered as trapped species under kinetic control, respectively. In general, self-sorting events are directed by the same intermolecular forces which govern any molecular recognition process, *i.e.* H-bonds,[154] metal-ligand interactions,[155] electrostatic interactions,[156] π-π stacking,[157] and solvophobic effects;[158] and therefore, the factors that determine these recognition events will compromise the fidelity of the self-sorting processes.

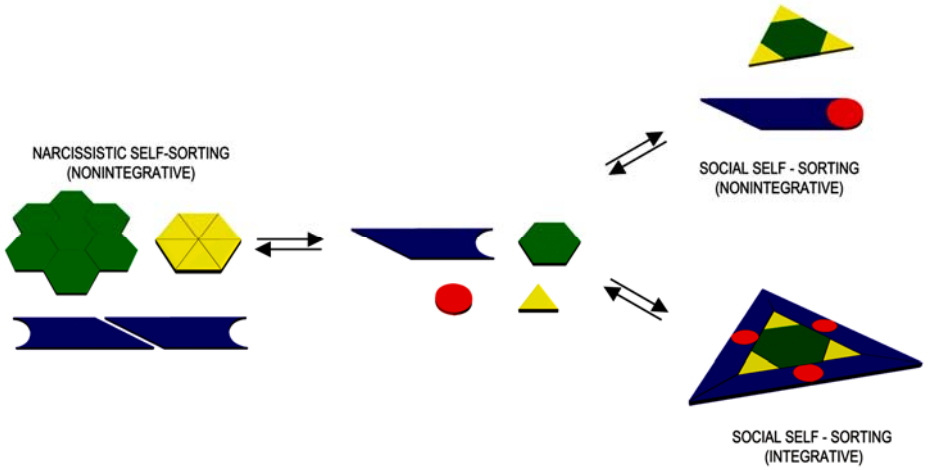


Figure 40. Schematic representation of the different types of self-sorting.[153]

[150] A. X. Wu, L. Isaacs, *J. Am. Chem. Soc.* **2003**, *125*, 4831–4835.
[151] A. Shivanyuk, J. Rebek, Jr., *J. Am. Chem. Soc.* **2002**, *124*, 12074–12075.
[152] P. N. Taylor, H. L. Anderson, *J. Am. Chem. Soc.* **1999**, *121*, 11538–11545.
[153] M. M. Safont-Sampere, G. Fernández, F. Wüthner, *Chem. Rev.* **2011**, *111*, 5784–5814.
[154] a) K. A. Jolliffe, P. Timmerman, D. N. Reinhoudt, *Angew. Chem. Int. Ed.* **1999**, *38*, 933–937; b) P. S. Corbin, L. J. Lawless, Z. T. Li, Y. G. Ma, M. J. Witmer, S. C. Zimmerman, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5099–5104; c) Y. G. Ma, S. V. Kolotuchin, S. C. Zimmerman, *J. Am. Chem. Soc.* **2002**, *124*, 13757–13769.
[155] a) D. L. Caulder, K. N. Raymond, *Angew. Chem. Int. Ed.* **1997**, *36*, 1440–1442; b) E. J. Enemark, T. D. P. Stack, *Angew. Chem. Int. Ed.* **1998**, *37*, 932–935; c) R. Stiller, J.-M. Lehn, *Eur. J. Inorg. Chem.* **1998**, 977–982; d) M. Albrecht, M. Schneider, H. Röttele, *Angew. Chem. Int. Ed.* **1999**, *38*, 557–559; e) D. Schultz, J. R. Nitschke, *Angew. Chem. Int. Ed.* **2006**, *45*, 2453–2456; f) Y. M. Legrand, A. van der Lee, M. Barboiu, *Inorg. Chem.* **2007**, *46*, 9540–9547; g) M. Barboiu, F. Dumitru, Y. M. Legrand, E. Petit, A. van der Lee, *Chem. Commun.* **2009**, 2192–2194; h) J. R. Nitschke, *Acc. Chem. Res.* **2006**, *40*, 103–112.
[156] W. Jiang, H. D. F. Winkler, C. A. Schalley, *J. Am. Chem. Soc.* **2008**, *130*, 13852–13853.
[157] A. D. Shaller, W. Wang, H. Y. Gan, A. D. Q. Li, *Angew. Chem. Int. Ed.* **2008**, *47*, 7705–7709.
[158] a) B. Bilgicer, X. Xing, K. Kumar, *J. Am. Chem. Soc.* **2001**, *123*, 11815–11816; b) N. A. Schnarr, A. J. Kennan, *J. Am. Chem. Soc.* **2003**, *125*, 667–671.

This phenomenon is already demonstrated in DNA and RNA where, according to the Watson-Crick base pairing model, guanosine binds specifically to cytidine and cannot strongly bind to thymidine/uridine or adenosine. As we already know, the nucleobase pair G–C forms a *DAA–ADD* union, but 2-aminoadenine and uracil form an *ADA–DAD* union. In order to exploit this phenomenon, self-sorting experiments (1D and 2D ^{1}H NMR, as well as absorption and emission spectroscopy) have been carried out at the level of the cyclic tetramer in our research group. By self-sorting we understand a system where GC and AU species do not interact with each other in a mixture of the two but, instead, associate independently. The ^{1}H NMR spectrum of the mixture of **GC$_{rib1}$** and **AU$_{rib1}$** cyclic tetramers showed how the characteristic proton signals of each species were not affected by the presence of the other (Figure 41a). NOESY experiments also revealed no interaction (Figure 41b). More interestingly, similar experiments showed that **GC$_{rib1}$** and **iGiC$_{rib1}$** associate separately, as well (Figure 41c,d), despite having both nucleobase pairs a compatible asymmetric H-bonding pattern. In fact, even if the binding between the G and iC or iG and C derivatives is possible and of related strength,[148] the angle formed is 210^{o} and prevents the macrocyclization process.[159] Therefore, self-sorting in the GC+iGiC system is basically ruled by the strong chelate cooperativity shown by both monomers, which results in strong *EM* values when **GC$_{rib1}$** and **iGiC$_{rib1}$** cyclotetrimerize independently.

[159] Unpublished results.

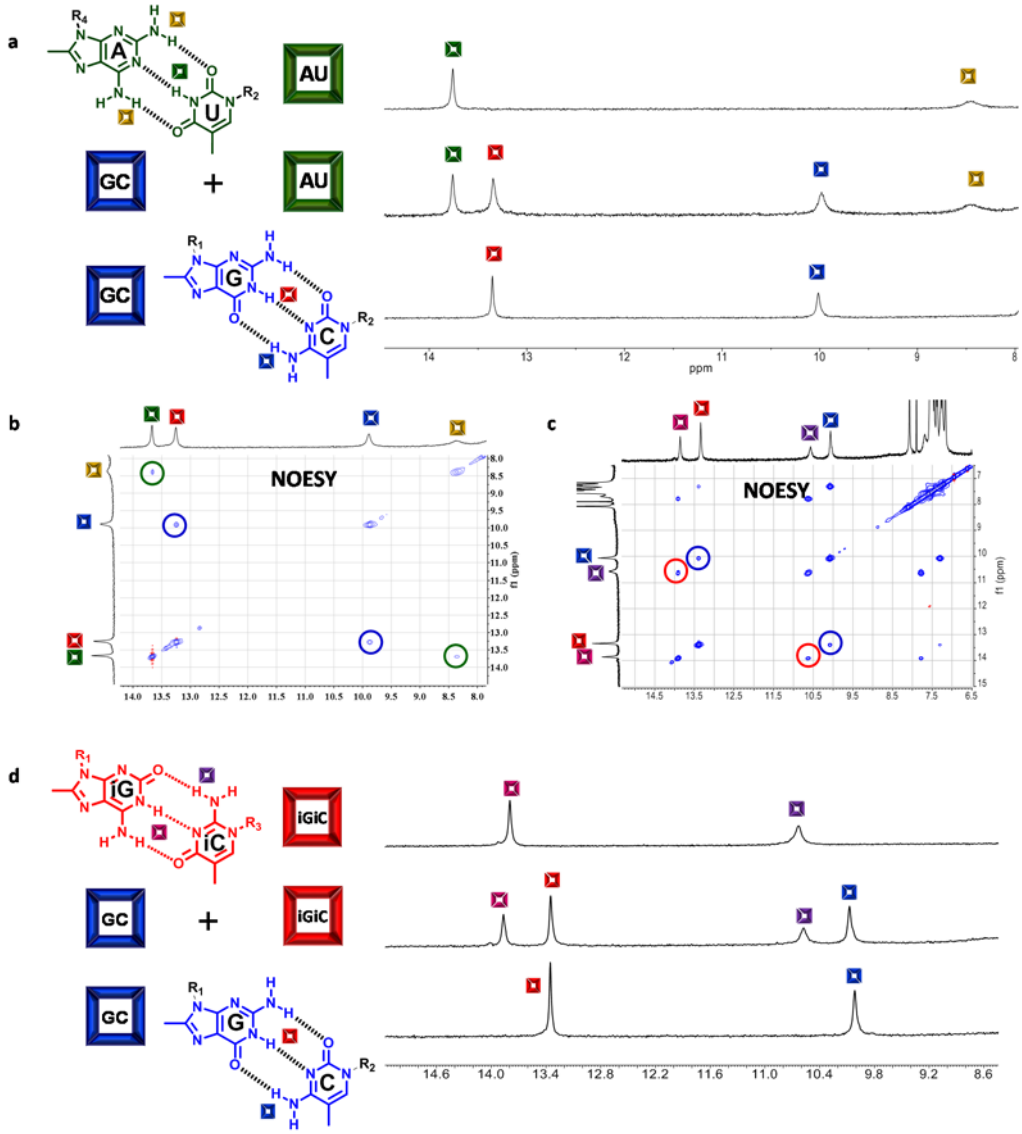


Figure 41. ^{1}H NMR and NOESY self-sorting experiments between GC_{rib1}, AU_{rib1} and $iGiC_{rib1}$ cyclic tetramers in $CDCl_3$. (**a**) ^{1}H NMR spectrum showing that the chemical shifts of the proton signals of $\mathbf{GC_{rib1}}$ and $\mathbf{AU_{rib1}}$ cyclic tetramers are not affected when the two species are mixed (middle spectrum). (**b**,**c**) The NOESY experiments of the mixtures of $\mathbf{GC_{rib1}}$ + $\mathbf{AU_{rib1}}$ and $\mathbf{GC_{rib1}}$ + $\mathbf{iGiC_{rib1}}$ cyclic tetramers show no correlation between the characteristic H-bonded proton signals of the nucleobase derivatives. (**d**) ^{1}H NMR spectrum showing that $\mathbf{GC_{rib1}}$ and $\mathbf{iGiC_{rib1}}$ cyclic tetramers associate independently when mixed.

Additionally, emission spectroscopy experiments in toluene between donor-acceptor (*D*-*A*) FRET couples afforded interesting results. When mixing $\mathbf{GC_{rib2}}$, carrying a benzodithiophene (*D*) as central block with monomer $\mathbf{GC_{rib3}}$, containing a BODIPY derivative (*A*), a set of spectra, showing a gradual decrease of the donor emission at the expense of

an increase of the acceptor emission, were obtained over a period of 20 hours (Figure 42a). These changes are indicative of an evolution of the initial situation to a statistical mixture of cyclic tetramers at equilibrium *via* monomer exchange. In some of these new tetramers formed, donor and acceptor monomers coexist, and thus energy transfer takes place. However, when mixing **GC$_{rib2}$** with **AU$_{rib2}$**, substituted with the BODIPY central block (*D*), no evolution was detected whatsoever: the system self-sorts (Figure 42b).[160]

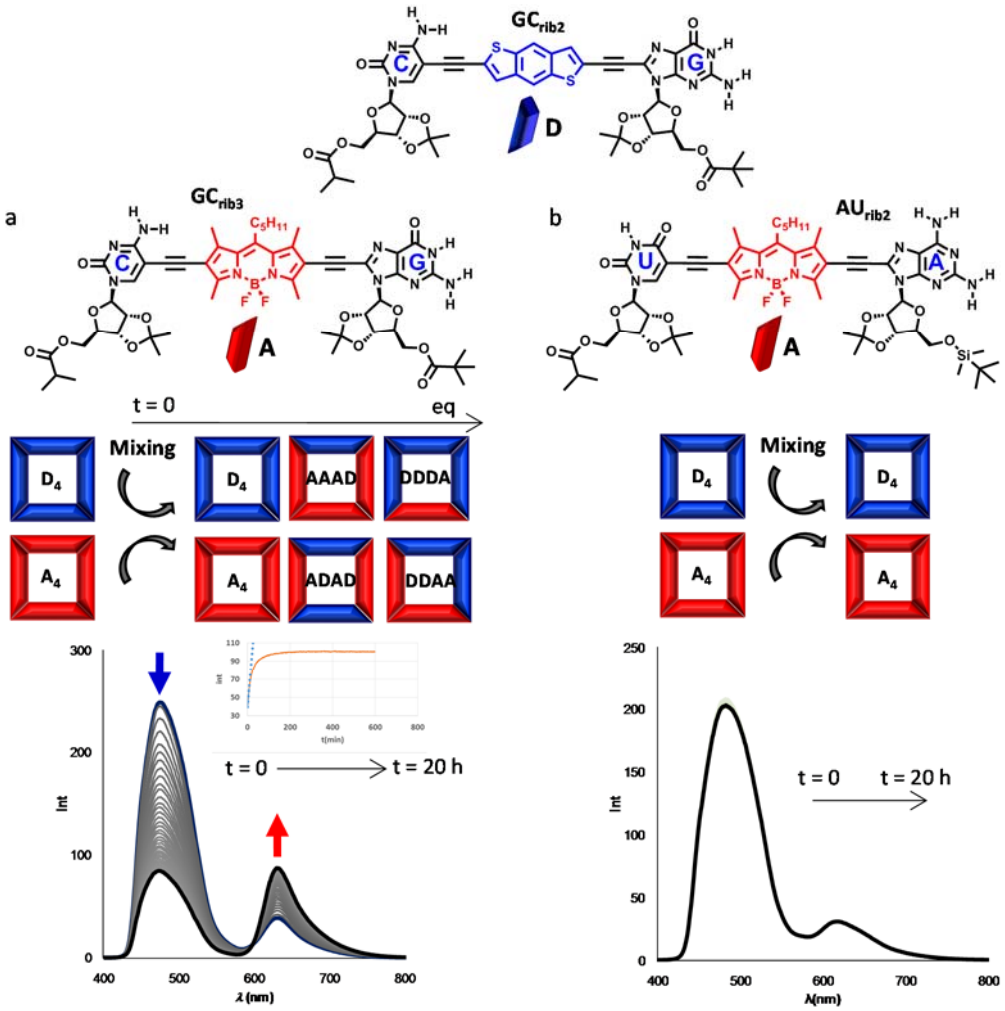


Figure 42. **Mixing experiments of donor (benzodithiophene) and acceptor (BODIPY) monomers in toluene.** (**a**) Mixing **GC$_{rib2}$** and **GC$_{rib3}$** results in emission spectra changes over a period of 20 hours, which is indicative of the evolution of the initial situation to a statistical mixture of cyclic tetramers at equilibrium. (**b**) Mixing **GC$_{rib2}$** and **AU$_{rib1}$** leads in contrast to no spectral change over time, indicating that the initial and final situations are identical.

[160] Unpublished results.

5) Multicomponent Cyclic Systems.

The true structural versatility of our system comes from the possibility of forming multicomponent cyclic systems (Figure 43). By covalently coupling different nucleobase derivatives, targeted dinucleoside monomers can be synthesized in a straightforward fashion and, by strategic combination of these monomers, a wide diversity of multicomponent systems can be easily assembled.

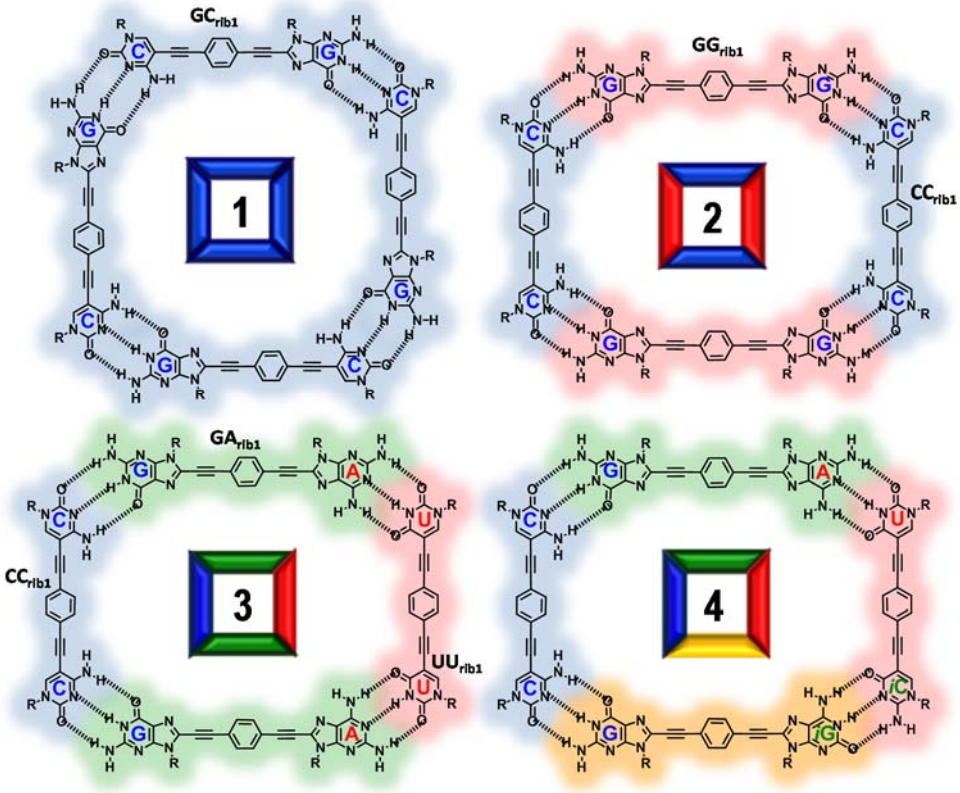


Figure 43. Multicomponent systems. Examples of cyclic tetramers constituted by one (A_4), two (ABAB), three (ABAC) and four (ABCD) components.

Two-component (ABAB) systems can be built from the combination of two monomers having complementary bases at the edges. For instance, the ABAB system formed by symmetric **GG$_{rib1}$** and **CC$_{rib1}$** monomers has been thoroughly investigated in our group (Figure 43, top right corner), although other combinations are possible.[161] It is worth to note that the shape of the cyclic tetramer can be regulated in these systems by increasing the length of one of the monomers. The formation of the **GG$_{rib1}$-CC$_{rib1}$** ABAB system was

[161] Unpublished results.

first investigated by ^{1}H NMR (Figure 44). Two main aspects should be underlined when comparing this system with the previously studied **GC$_{rib1}$** one-component cyclic tetramer: (1) a much lower thermodynamic stability, deriving from smaller *EM* values; and (2) a fast exchange between the monomer constituents in the NMR timescale. This was readily demonstrated in titration experiments in $CDCl_3$. Upon addition of increasing amounts of **CC$_{rib1}$** to **GG$_{rib1}$**, the **GG$_{rib1}$** amide proton signal experienced a downfield shift up to 13.2 ppm after reaching 1:1 stoichiometry, which is indicative of G-C binding (Figure 44a). On the other hand, a Job plot was performed in order to determine the stoichiometry of the binding event and the obtained results confirmed a 1:1 binding (Figure 44b).

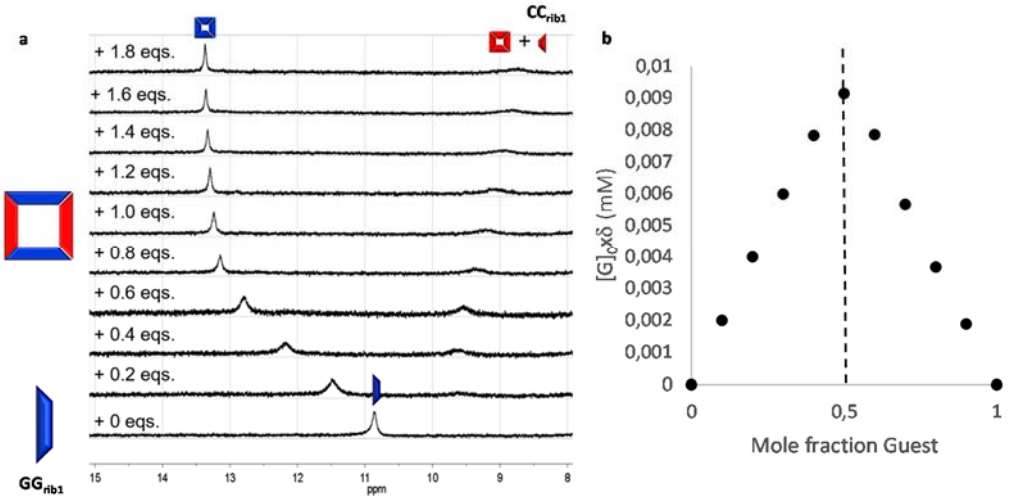


Figure 44. **Study of the two-molecule ABAB system formed by GG$_{rib1}$ and CC$_{rib1}$.** (**a**) Titration experiment of **GG$_{rib1}$** (C = 10^{-2} M) with **CC$_{rib1}$** in $CDCl_3$ (+1% DMSO-d_6) at 25 ^{o}C monitored by ^{1}H NMR. (**b**) Corresponding Job plot experiment monitored by ^{1}H NMR indicating the formation of a 1:1 complex.

In the same way, *three-component (ABAC)* systems can be built by combining 3 different monomers presenting four different nucleobases: for instance, **CC$_{rib1}$**, **UU$_{rib1}$**, and 2 equivalents of **GA$_{rib1}$** (Figure 43, bottom left corner). Three-component systems are very interesting to the group because they give the possibility of positioning different functional units at defined distances, for instance energy donors and acceptors, p- and n-type semiconducting molecules, etc. (see Project 2 in Figure 33).

Finally, the *four-component (ABCD)* tetramer is the most complex system and requires a third nucleobase pair, such as the iG-iC couple (Figure 43, bottom right corner). This system has to be designed very carefully to avoid the formation of other tetramers and is currently under study in our research group.

Objectives.

Despite the great effort undertaken by the scientific community to precisely tailor materials at the macroscopic scale, the rigorous control and transmission of the molecular properties at the nanoscale level, still remains a challenging task. A contribution in the field of supramolecular chemistry and self-assembly for the preparation of finely-tuned nanostructures on different surfaces will be pursued in this Thesis. Taking natural examples as models and DNA nucleobase derivatives as pawns, we will try to go further in the utilization of such extraordinary selective and directional units and explore molecular recognition on the surface. In this Thesis, *we aim to translate the supramolecular chemistry developed in solution by our research group to the formation of 2D networks*. More specifically, we would like to ***develop an unconventional and versatile strategy based on molecular self-assembly toward nanostructured porous surfaces able to specifically recognize guest molecules as a function of their size, shape and chemical structure***.

The self-assembly strategy includes the first stage explained above, where specific *cyclic tetramers* are formed from four monomeric subunits by H-bonding interactions between *directors* (*i.e.* self-complementary H-bonded nucleobases). These discrete tetramers then assemble into ordered porous networks guided by the cooperative action of different interactions (mainly H-bonding, van der Waals forces and interactions with the substrate) (Figure 45).

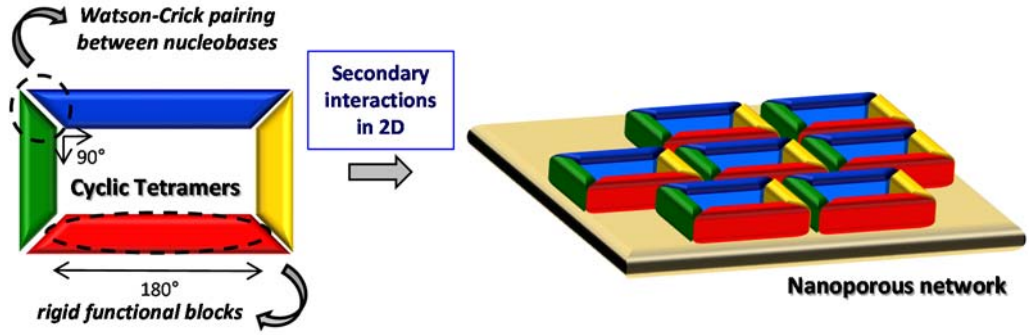


Figure 45. Self-assembly strategy toward nanostructured self-assembled nanoporous networks.

In order to fulfill this main objective, various goals have been addressed, gradually increasing the degree of complexity (Figure 46). Four approaches, along with different experimental techniques, have been considered, each corresponding to one of the Chapters defining this Thesis. In *Chapters 2 and 3*, the self-assembly process has been carried out on surfaces, and studied by STM at the solid-liquid interface and under UHV, respectively. A third approach, described in *Chapter 4*, consisted in the preparation of Langmuir-Blodgett films, where monolayers are first assembled at the water-air interface and then transferred to dipping substrates. This strategy offers the additional possibility of

controlling the formation of multilayered porous networks. In *Chapter 5*, covalently fused-monomers are targeted with a dual objective: to guide 2D self-assembly solely relying on strong and directional Watson-Crick interactions; and to promote the formation of H-bonded organic frameworks (HOFs) by vertical growth of these highly ordered monolayers into crystalline 3D materials.

The first objective of this Thesis (*Chapter 1*) is, on the other hand, purely synthetic: the design and synthesis of the molecular components that will constitute our target monomers. We aim to prepare a wide family of nucleobase derivatives substituted at one end with an ethynyl moiety, as well as a complete variety of dihalogenated building blocks to incorporate as the core of the targeted monomers. We will explore the reactivity of DNA-nucleobase derivatives and optimize synthetic routes toward halogenated and ethynylated derivatives, appropriately functionalized at the **R^1** position (see Figure 34). The chemistry of π-conjugated flat aromatic molecules that could serve as central blocks will also be investigated. Finally, a convergent coupling route toward the targeted monomers comprising a chosen central block and suitable self-assembling units at the corners will be pursued.

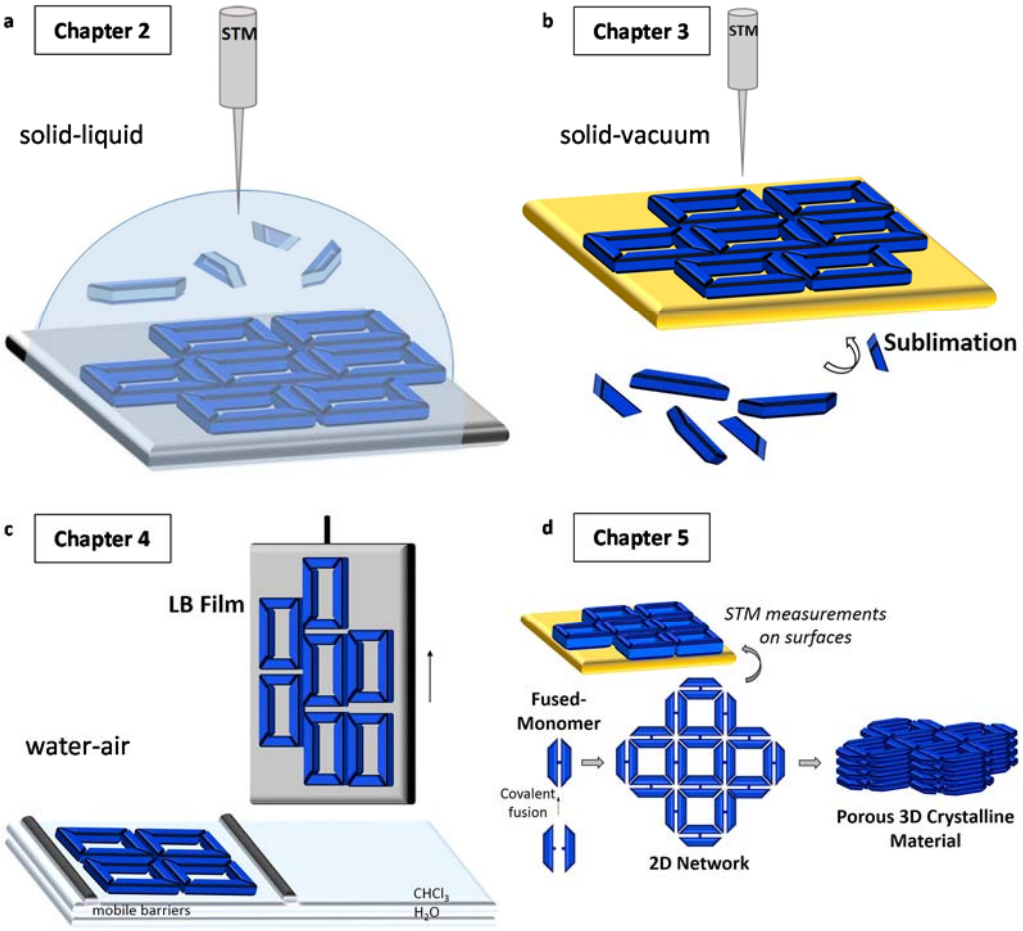


Figure 46. **Schematic representation of the self-assembly and imaging strategies approached in this Thesis:** (**a**) 2D self-assembly at the solid-liquid interface *via* drop-casting; (**b**) 2D self-assembly at the solid-vacuum interface *via* sublimation; (**c**) 2D self-assembly at the water-air interface *via* the preparation of Langmuir-Blodgett films; (**d**) 2D and 3D self-assembly of fused-monomers into H-bonded organic frameworks.

The second objective would be the study of the self-assembly process on the surface, where association toward cyclic tetramers, instead of open oligomeric structures or other kind of networks, will be pursued. For this purpose, a series of ditopic molecules, carrying complementary nucleobase derivatives on each side, will be prepared and studied. This will mainly be investigated by STM, from conditions that afford high resolution images (at vacuum) to a more realistic environment (ambient conditions). This objective will principally be developed in *Chapters 2 and 3*.

In *Chapter 2*, we aim to study the self-assembly of different monomers onto HOPG substrates by STM analysis at the solid-liquid interface. The objective would be to stabilize cyclic tetramer networks on the surface that would lead to rationally nanostructured

porous surfaces. We want to accomplish a series of sub-objectives that have been listed below (Figure 47), in which we will gradually increase the degree of supramolecular control and thus the complexity of the network structure. Compared to STM under high vacuum, where the deposited molecules have to be light enough in order to sublimate them, the STM at the solid-liquid interface, where the molecules are drop-casted from solution, offers greater versatility, a higher degree of dynamics, and grants more freedom to study these various topics.

Optimizing cyclic networks. When concentrated on the surface, the intramolecular event that shifts the equilibria toward cyclization do not necessarily apply and closed H-bonded species might be in equilibrium with open oligomers, even if the molecule is preorganized for ring closure. Therefore, the first task will be to stabilize on the surface cyclic tetramers over polymeric species. Then, the secondary interactions that stabilize the 2D network will be analyzed. Finally, inner and outer cyclic tetramer cavities will be identified in order to fill the unspecific space left between macrocycles.

Host-guest. Since 2D porous networks are targeted, the anticipated ability of the nanostructured surfaces to host appropriate guest molecules will be studied. The objective is to self-assemble a persistent nanoporous densely-packed network of cyclic tetramers onto the surface and then, find a suitable guest to fill selectively the pores.

Surface chirality. An effort will be made to gain insight into how we can address chirality control on the surface. We wish to go a step forward in controlling the chirality of our system in 2D and different approaches to rule over pore chirality on the surface will be developed.

Size discrimination. Once the viability of a host-guest system has been checked, the ability of the 2D H-bonded porous networks to host different guest molecules as a function of their size will be studied.

Functionality. Since STM images can be considered as a map of electronic density, we intend to explore the effect, on the actual STM images, of the presence of different functional π-electron-rich/-poor blocks in the monomers. Our versatile supramolecular approach, in which multiple kinds of multicomponent macrocycles can in principle be accessed, could ultimately become a potent tool to hold different π-functional molecules at specific positions on a substrate.

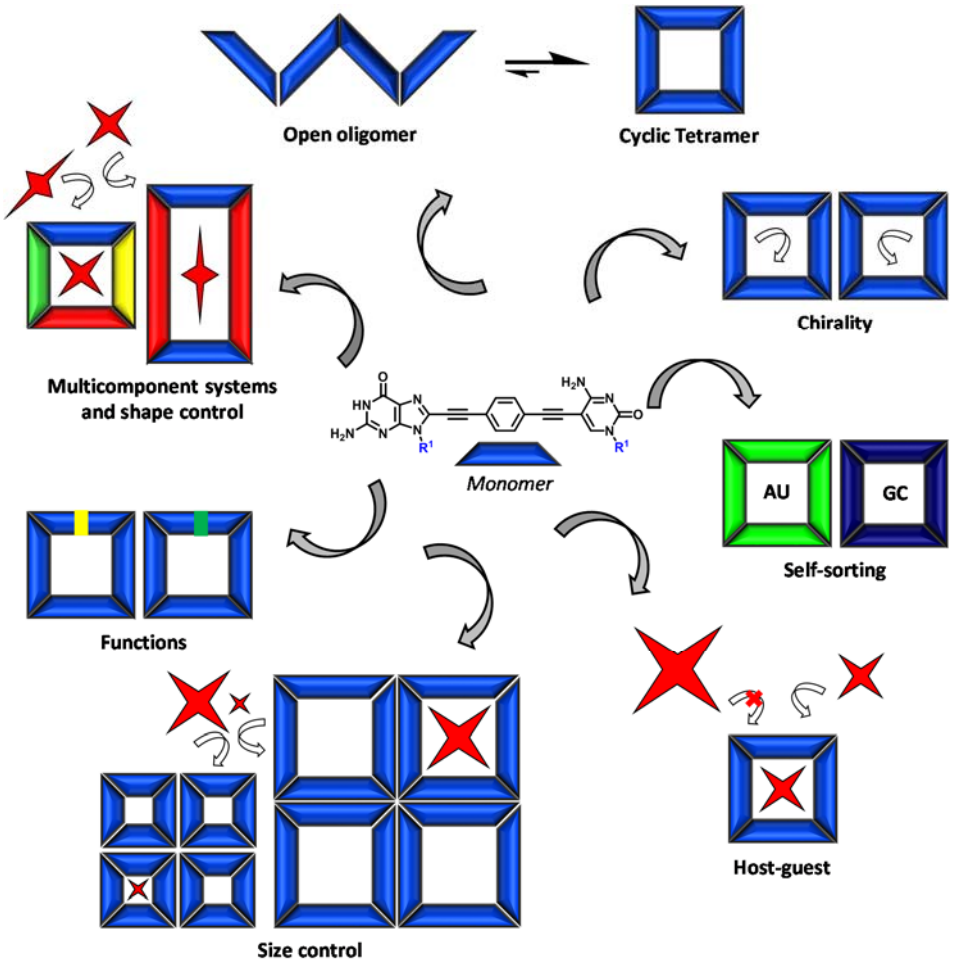


Figure 47. Molecular recognition on nanostructured surfaces: schematic view of different degrees of supramolecular control that we aim to achieve in this Thesis.

Self-sorting. Again, the versatility of our supramolecular toolkit wants to be profited by studying in depth the self-assembly process on the surface of, not only GC monomers, but also AU species. The specificity and complementarity of the H-bond between nucleobases will be analyzed in order to go a step beyond the state-of-the-art in molecular recognition on surfaces. Self-sorting experiments will be carried out as proof of concept of DNA-nucleobase complementarity. By self-sorting phenomena on the surface we understand a system where GC and AU monomers are co-deposited on the substrate but the domains of cyclic tetramers are not mixed and remain formed by only one component.

Shape discrimination. Once enough insight has been gained into how to form robust 2D porous networks formed by one-component cyclic tetramers, two-component macrocycles will be assembled on the surface. Since monomers of different lengths can be associated,

not only square-shaped tetramers, but also rectangular-shaped cyclic species will be approached in this way. Additionally, the ability of the 2D porous surfaces to host different guest molecules as a function of their shape will be explored.

In *Chapter 3* we will study the deposition by sublimation of lighter monomers and the formation of organized networks on diverse metal surfaces under vacuum. The monolayers will be analyzed by UHV-STM, profiting from the high resolution of this technique. Additionally, theoretical calculations will be carried out for a deeper comprehension of the structure of the self-assembled monolayer.

In order to give a first step toward a hypothetical technological utilization, the possibility to allow the 2D networks to grow on the third dimension has to be considered. In *Chapter 4*, the preparation of Langmuir-Blodgett (LB) films based on our monomers will be addressed. Porous 2D networks of cyclic tetramers, similar to the ones targeted previously in Chapters 2 and 3, will be first self-assembled at the water-air interface. Then, these H-bonded Langmuir films will be transferred by the vertical deposition method to various solid substrates. Finally, the LB films will be studied by different scanning probe microscopy techniques. We ambition accessing, in a controlled manner, materials with a given number of layers (bilayer, trilayer and so on).

As a final objective of this Thesis (*Chapter 5*), we targeted to bring this supramolecular chemistry to the third dimension in crystalline materials. By letting the number of layers grow along the "*z*" axis, we intend to form supramolecular organic frameworks (SOFs) similar to MOF or covalent organic framework (COF) materials. However, Watson-Crick H-bonding will be here central in leading the organization of the 2D self-assembled layers.[162] These systems will be based on covalently bound double monomers or "fused-monomers". With them, we intend to reach more robust and densely-packed 2D networks that, under the appropriate conditions, will stack and guide the formation of 3D porous frameworks. First of all, a theoretical approximation of these organic materials will be considered through the construction of different molecular models, where the inner and outer pores of the SOF material will be calculated. Particular attention will be paid in these models to filling any non-selective void space left between cyclic tetramers, whose cavity will be responsible for the potential hosting/adsorption properties of the future nanoporous materials. Once a suitable design is found, the synthesis of the corresponding fused-monomers will be approached and their potential to self-organize in H-bonded organic frameworks (HOFs) will be tested.

[162] a) W. Yang, A. Greenaway, X. Lin, R. Matsuda, A. J. Blake, C. Wilson, W. Lewis, P. Hubberstey, S. Kitagawa, N. R. Champness, M. Schröder, *J. Am. Chem. Soc.* **2010**, *132*, 14457–14469; b) J. Tian, P. K. Thallapally, B. P. McGrail, *CrystEngComm* **2012**, *14*, 1909–1919.

Chapter 1.

Design and Synthesis of the Molecular Components.

1.1. Design of the monomer components.

The key structural elements that afford self-assembly in cyclic tetramers have been described before in the Background Section. In this Thesis we will keep such design and therefore, the target monomers share a common structure based on rigid and planar π-conjugated *central blocks* that are linearly substituted at both ends with nucleobase directors, which will guide self-assembly via H-bonding in 90º angles (Figure 48).

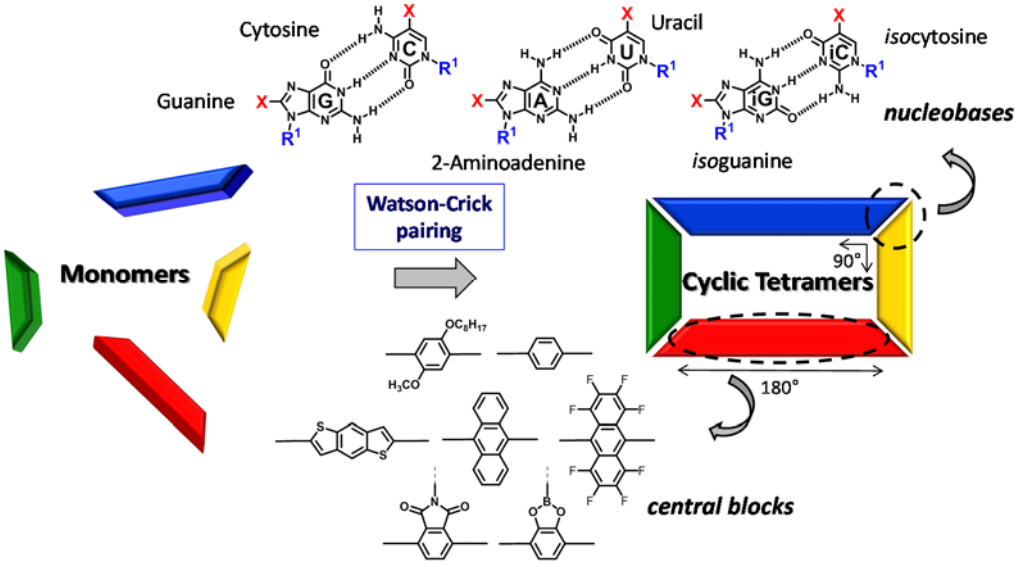


Figure 48. The self-assembly strategy.

The structure and preparation of each molecular component, nucleobase directors and central blocks, will be explained separately in this first Chapter. The combination of these fragments to yield the target monomers for each application will be correspondingly explained in the subsequent Chapters.

1.1.1. Central Blocks.

Central-block molecules have been carefully designed taking into account different issues and requirements: (1) linear substitution at both termini; (2) the possibility of having specific lateral groups; (3) electronic properties, since the electron density of the central blocks can also be used to our profit; and (4) length, which will regulate cyclic tetramer size. Numerous π-conjugated systems can fulfill such characteristics as central blocks in the

monomer and can provide our molecular targets with characteristic functionalities and meaningful potential technological applications.[163]

Therefore, the nature of the central block is essential for our purposes. Many of the target assemblies will be studied by STM, where the enhanced tunneling current associated with some physisorbed electron-dense moieties is translated into a characteristic brightness. In that way, the STM images of the adlayer will present a specific contrast, useful for image interpretation.[164] For the deposition on top of planar substrates, some affinity for the chosen substrate needs to be taken into account. Therefore, molecules with a large π-conjugated area can be adopted (such as polycyclic aromatic compounds or oligo-*p*-phenylene derivatives substituted with long alkyl chains), whose affinity for a graphite substrate (see Chapter 2), for instance, is well-known.[165] Other parameters such as low molecular weight for sublimation purposes (see Chapter 3) have to be considered as well. In Chapter 5, molecules formed out of two covalently fused-monomers will be pursued. Such monomer fusion will be made irreversible, *via* phthalimide linkers; or reversible, through boronate ester formation, a reaction that has been successfully developed for the construction of COF materials.[166] Therefore, phthalic anhydride and catechol central blocks have also been targeted.

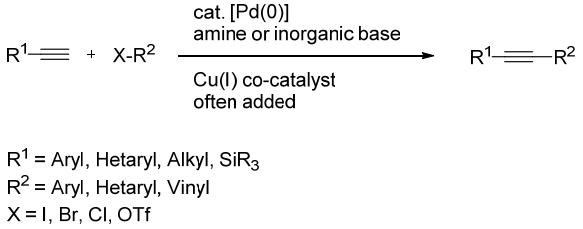
1.1.2. Nucleobase Directors.

The moieties located at the end caps of the central blocks are responsible on their own for the formation of sophisticated supramolecular architectures. As stated in the Background Section, naturally occurring DNA bases and synthetic nucleobase analogues (G, C, A, U, iG and iC; Figure 48), equipped with an array of H-bonding donor and acceptor groups, have been chosen as self-assembling directors.[20c,51d,52,147a,167] Since the main objective of this Thesis is to study two-dimensional assemblies and networks, planar substituents, such as alkyl chains, have been incorporated at the **R^1** position of the nucleobase derivatives instead of the bulky riboses, present in the nucleosides previously studied in the Group. These groups are more appropriate for the adsorption onto surfaces, endow the molecules with sufficient solubility and can, additionally, serve to fill unspecific secondary cavities of the 2D-network.

[163] a) T. Yamamoto, T. Fukushima, T. Aida, *Adv. Polym. Sci.* **2008**, *220*, 1–27; b) S. Hecht, A. Khan, *Angew. Chem. Int. Ed.* **2003**, *42*, 6021–6024; (c) M. Barboiu, A. Cazacu, M. Michau, R. Caraballo, C. Arnal-Herault, A. Pasc-Banu, *Chem. Eng. Process.* **2008**, *47*, 1044–1052.
[164] a) J. K. Spong, H. A. Mizes, L. J. LaComb Jr, M. M. Dovek, J. E. J. S. Foster, *Nature* **1989**, *338*, 137–139; b) D. M. Cyr, B. Venkataraman, G. W. Flynn, *Chem. Mater.* **1996**, *8*, 1600–1615; c) L. C. Giancarlo, G. W. Flyn, *Annu. Rev. Phys. Chem.* **1998**, *49*, 297–336.
[165] A. J. Groscek, *Proc. R. Soc. Lond. A* **1970**, *314*, 473–498.
[166] A. P. Côté, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger, O. M. Yaghi, *Science* **2005**, *310*, 1166–1170.
[167] a) S Sivakova, S. J. Rowan, *Chem. Soc. Rev.* **2005**, *34*, 9–21; b) J. L. Sessler, M. Sathiosatham, C. T. Brown, T. A. Rhodes, G. Wiederrecht, *J. Am. Chem. Soc.* **2001**, *123*, 3655–3660.

1.1.3. Cross-Coupling between Directors and Central Blocks: The Sonogashira Reaction.

Sonogashira coupling reactions[168] will be extensively used to attach the different components of our target monomers: self-assembling directors and central blocks. Nowadays, coupling reactions catalyzed by palladium are among the most important procedures for sp^2-sp carbon-carbon bond formation in organic synthesis and Sonogashira cross-coupling reactions are frequently employed in the synthesis of natural products, heterocycles, molecular electronics, conjugated polymers or nanostructures; and many examples are available in the literature.[169] A Sonogashira coupling refers to the Pd-catalysed C-C bond formation process which is able to bond a terminal sp hybridized carbon from an alkyne with a sp^2 carbon of an aryl vinyl halide (or triflate). The japanese researchers Sonogashira, Tohda and Hagihara proved in 1975 how this process could be easily performed at room temperature using a palladium source, such as $Pd(PPh_3)_2Cl_2$ as catalyst, combined with a co-catalytic amount of CuI in an amine solvent (Scheme 2).[170]



Scheme 2. Scheme for a general Sonogashira Cross-Coupling reaction.

The copper-mediated transmetalation of alkynes is combined with the catalytic properties of palladium. It is however necessary to deoxygenate the solvent of the reaction in order to avoid unwanted alkyne homocoupling formation through a copper-mediated Hay/Glaser reaction.[171] At present day, the mechanism of this reaction is supposed to take place through two independent catalytic cycles (Figure 49).[168b]

[168] a) R. Chinchilla, C. Nájera, *Chem. Soc. Rev.* **2011**, *40*, 5084–5121; b) R. Chinchilla, C. Nájera, *Chem. Rev.* **2007**, *107*, 874–922; c) H. Doucet, J.-C. Hierso, *Angew. Chem. Int. Ed.* **2007**, *46*, 834–871.
[169] a) L. A. Agrofoglio, I. Gillaizeau, Y. Saito, *Chem. Rev.* **2003**, *103*, 1875–1916; b) R. Bielski, Z. J. Witczak, *Chem. Rev.* **2013**, *113*, 2205–2243.
[170] K. Sonogashira, Y. Tohda, N. Hagihara, *Tetrahedron Lett.* **1975**, *16*, 4467–4470.
[171] G. Evano, N. Blanchard, M. Toumi, *Chem. Rev.* **2008**, *108*, 3054–3131.

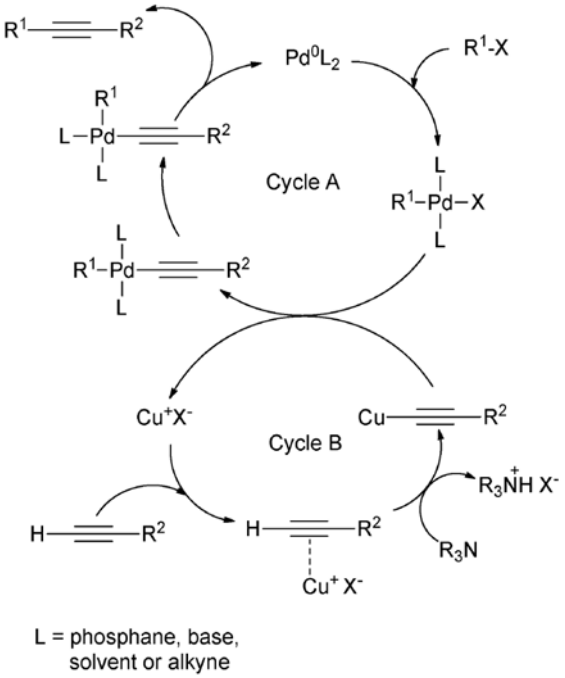


Figure 49. Supposed mechanism for the copper-cocatalysed Sonogashira reaction.[168a]

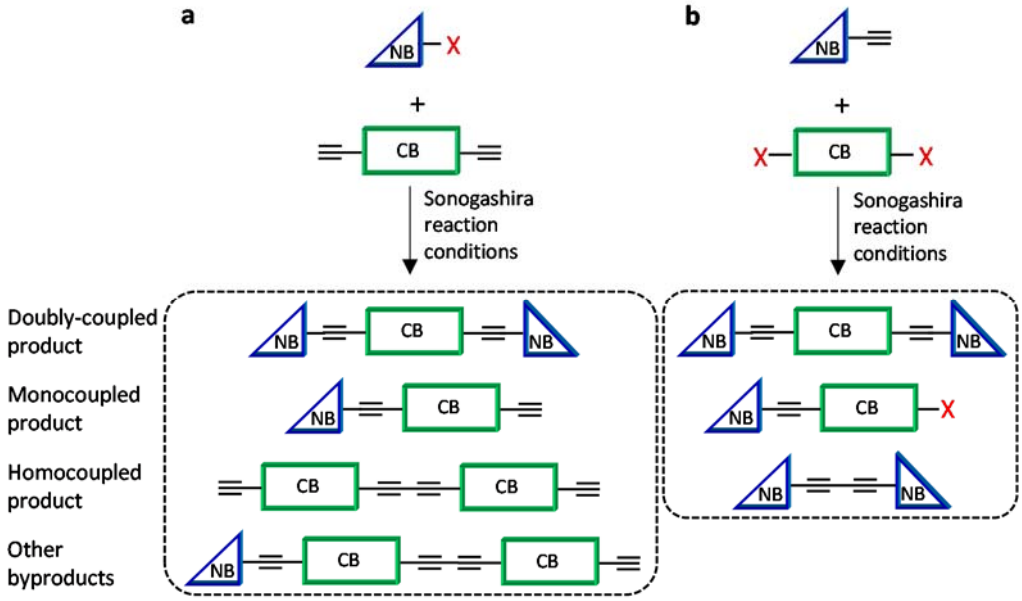
The first "palladium-cycle" (cycle A) starts in the catalytically active species Pd^0L_2, which can be formed from Pd^0 complexes such as $Pd(PPh_3)_4$, or can be created from Pd^{II} complexes such as $Pd(PPh_3)_2Cl_2$, through formation of a $[Pd^{II}L_2(C{\equiv}CR^2)_2]$ species. The latter gives $[Pd^0L_2]$ after reductive elimination by forming a diyne. Amines may also reduce Pd^{II} to Pd^0 through formation of iminium cations.[172] Once complex $[Pd^0L_2]$ has been formed, the first step in the catalytic cycle is initiated by oxidative addition of the aryl or vinyl halide, which is considered to be the rate-limiting step of the Sonogashira reaction. The barriers of oxidative addition of ArX (X = Cl, Br, I) increase in the order of ArI < ArBr < ArCl, as it is believed to be an electron-donating step.[173] The formed $[Pd^{II}R^1L_2X]$ adduct is then transformed into a $[Pd^{II}L_2R^1(C{\equiv}CR^2)]$ species after transmetalation with a copper acetylide formed in the "copper-cycle" (cycle B). This adduct suffers reductive elimination to the final alkyne, regenerating $[Pd^0L_2]$.

In the presence of an oxidant such as molecular oxygen, Pd^0 complexes can be oxidized in order to be incorporated in the catalytic cycle. The alkyne species is consumed and the Sonogashira coupling reaction can progress with the formation of a homocoupling byproduct.[171] To avoid this problem, deoxygenated solvents through freeze-pump-thaw

[172] H. Li, G. A. Grasa, T. J. Colacot, *Org. Lett.* **2010**, *12*, 3332–3335.
[173] C. Gottardo, T. M. Kraft, M. S. Hossain, P. V. Zawada, H. M. Muchall, *Can. J. Chem.* **2008**, *86*, 410–415.

cycles must be used. However small amounts of homocoupling side-products are still often observed.

Two possibilities arise for the coupling of central blocks and nucleobases (Scheme 3) and at the beginning of this Thesis we evaluated the most convenient route:



Scheme 3. Two possible combinations to carry out the Sonogashira coupling between central blocks (CB) and nucleobases (NB).

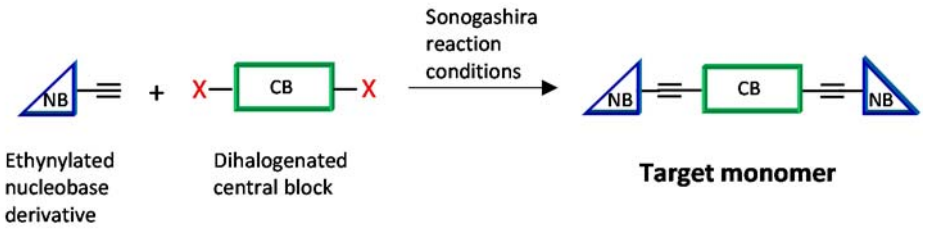
(**a**) The ethynyl spacers could, on the one hand, be first fused to the central building block. By following this route, three byproducts, in addition to the desired monomer, were obtained when coupling the central block functionalized with two triple bonds and the halogenated nucleobase. The isolated monocoupled byproduct could be used to reach unsymmetric monomers in a further step. The homocoupling byproduct that we observed presented two residues of the central block, separated by two ethynyl groups. A third byproduct, composed by this previous homocoupling product coupled to a nucleobase moiety, was also afforded by this route. It had a similar polarity than the targeted monomer and hindered its purification. None of those last two byproducts were useful to us.

(**b**) The ethynyl spacers could, on the other hand, be fused in a first step to the nucleobase. When carrying out the reaction between the dihalogenated block and the ethynyl-nucleobase, only two byproducts, other than the desired monomer, were identified. The mono-coupling product could serve for the synthesis of unsymmetric monomers, like in the previous route (**a**). The homocoupling byproduct corresponded to two nucleobases with two triple bonds between them. Even if no homocoupling product was in principle pursued, these dinucleobase derivatives may be of supramolecular interest.

After testing different reactions along this Thesis, route (**b**) was proven to be more convenient for our purposes. The route was more convergent and the valuable synthetic derivatives were spared in comparison to route (**a**). By following route (**b**), the separation of all the possible products was usually easier than the mixture afforded by route (**a**) and the obtained byproducts were sometimes useful to us. Therefore the inclusion of an ethynyl group at the 5-position (for the pyrimidines) or at the 8-position (in the case of the purines) will be targeted hereafter.

1.2. Synthesis of the monomer components.

The synthesis of the different halogenated functional central blocks, as well as the ethynylated self-assembling nucleobases, will be separately explained in this section. The final coupling between these components to generate the target monomers through consecutive Sonogashira reactions (Scheme 4) will be described in each of the following chapters.



Scheme 4. General scheme for the final Sonogashira coupling reaction affording the desired target monomers.

1.2.1. Central Blocks.

A series of seven dihalogenated aromatic molecules were considered as central blocks (Figure 50). Some of them are based on a simple benzene ring, others are polycyclic compounds and present a larger aromatic area. They form a complete and diverse family of blocks that will allow the study of different interesting parameters on the surface, such as substitution with alkyloxy chains, π-electronic area, electron density, as well as cyclic tetramer size. The optimized combination of these properties will be a useful tool for further objectives of this Thesis, as will be explained in each corresponding chapter. Some others will make possible the formation of more robust porous 2D supramolecular organic frameworks.

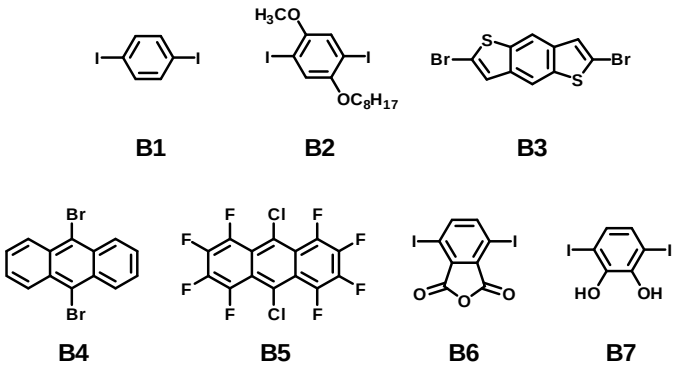
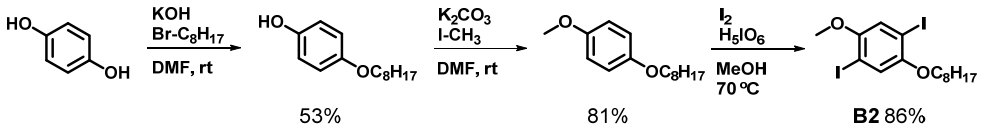


Figure 50. General structure of the targeted dihalogenated blocks **B1**-**B7**.

1,4-diiodobenzene (**B1**), 9,10-dibromoanthracene (**B3**) and 2,6-dibromobenzodithiophene (**B5**) were purchased from commercial suppliers. These three central blocks were chosen as they are planar, aromatic, π-conjugated molecules, easily available and they show a different but strong affinity for diverse substrates, through dispersion forces between electronic clouds. Furthermore these molecules are electron rich and they will provide our monomers with a specific label, easily recognizable in the STM images, which can be seen as electron density maps.

B2. 1,4-diiodo-2-methoxy-5-octyloxybenzene (**B2**) was synthesized from commercial 1,4-hydroquinone following the synthetic route shown in Scheme 5.



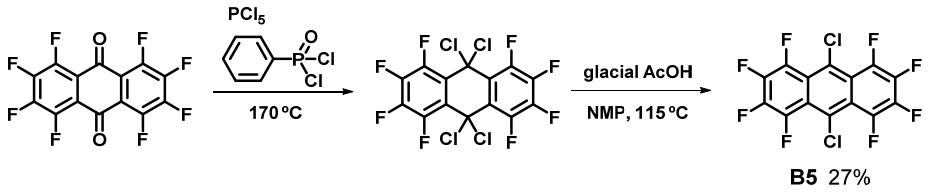
Scheme 5. Synthetic route to 1,4-diiodo-2-methoxy-5-octyloxybenzene (**B2**).

Unsymmetric substitution with alkyloxy chains has been incorporated to the phenyl ring of central block **B2**. The reasons for the choice of **B2** will be explained in Chapter 2. Building block **B2** was obtained through a three-step synthetic route, where the first reaction is the only yield-limiting one, as 1-hydroxy-4-octyloxybenzene had to be separated from a mixture containing the doubly-alkylated product. These reaction conditions were optimized by using a large excess of the starting material 1,4-hydroquinone in order to favor the formation of the mono-alkylated product. The second alkylation reaction can be carried out with a weaker base like potassium carbonate. 1-methoxy-4-octyloxybenzene was obtained in excellent yields as it is the only possible alkylated product and methyl iodide is a stronger electrophile than 1-bromooctane. Different iodination and bromination protocols and reagents were essayed for the synthesis of **B2** (iodine/iodic acid (I_2/HIO_3), *N*-iodosuccinimide (NIS), iodine monochloride (ICl), *N*-bromosuccinimide (NBS), bromine), but only a few worked properly. Strong acidic conditions (fuming sulfuric acid, acetic acid and iodine) afforded **B2** with a poor yield. However the use of the oxyacid with a higher oxidation state (periodic acid) in methanol[174] allowed obtaining 1,4-diiodo-2-methoxy-5-octyloxybenzene **B2** in excellent yields.

B5. Central block **B5** was considered because of its analogy with 9,10-dibromoanthracene (**B4**). Both have similar affinity for planar substrates as they share the same aromatic surface. However their electronic density is very different because fluorine atoms are extremely electronegative and they decrease the electron density of the polycyclic aromatic unit. Hence, similar molecules substituted with these related building blocks should appear very differently in the STM images because their HOMO and LUMO orbitals display a different electronic distribution within the monomer. This phenomenon could be

[174] S.-B. Ko, A.-N. Cho, M.-J. Kim, C.-R. Lee, N.-G. Park, *Dyes Pigm.* **2012**, *94*, 88–98.

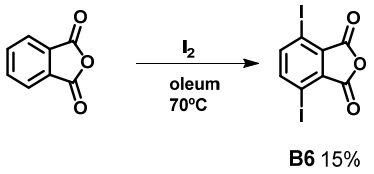
useful for future purposes (see Chapter 2) since it will help to recognize different molecules on top of the studied substrate. 9,10-dichlorooctafluoroanthracene (**B5**) was synthesized from commercial octafluoroanthraquinone as indicated in Scheme 6.[175]



Scheme 6. Synthetic route to 9,10-dichlorooctafluoroanthracene (**B5**).

Molecule **B5** was synthesized from commercial octafluoroanthraquinone in a single sequence without isolation of the tetrachlorinated intermediate. Octafluoroanthraquinone undergoes oxidative chlorination in the presence of phosphorous pentachloride and phenylphosphonic dichloride, which is more effective than phosphoryl chloride as it allows higher reaction temperatures due to its much higher boiling point. The final aromatization was induced by heating the crude to 115 ^{o}C in the presence of acetic acid. *N*-methyl-2-pyrrolidone (NMP) was used as an efficient aprotic solvent with high boiling point which is miscible with acetic acid and water.

B6. Different iodination protocols were essayed on commercial phthalic anhydride in order to obtain 3,6-diiodophthalic anhydride (**B6**). The mixture I_2/HIO_3 did not afford **B6**, nor did I_2/H_5IO_6. Also, the insolubility of the product, along with the ability to easily hydrolyze to the dicarboxylic acid, made the monitoring of the reaction, as well as its purification, tedious. However, strong acidic conditions (fuming sulfuric acid and iodine) afforded **B6** with low yields[176] after recrystallization with acetic acid (Scheme 7).

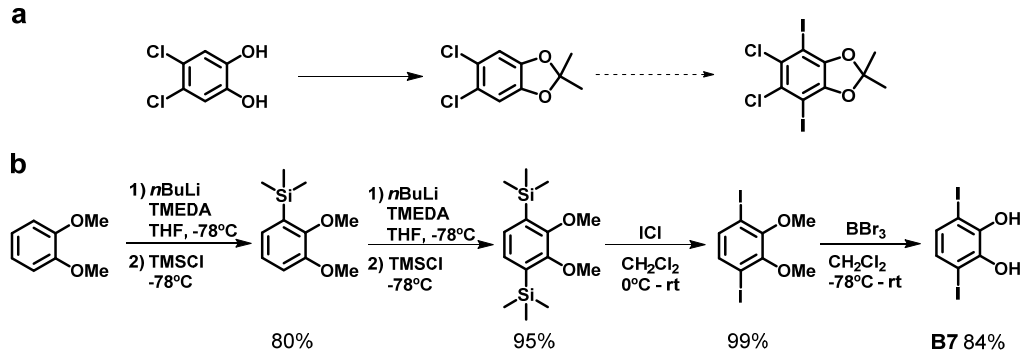


Scheme 7. Synthesis of 3,6-diiodophthalic anhydride (**B6**).

The low yield is due to occurrence of multiple mono-, di-, tri-, tetra-iodinated products that are very difficult to purify by recrystallization. Column chromatography could not be used as **B6** would hydrolyze to the dicarboxylic acid in the presence of silica gel, even deactivated with NEt_3.

[175] J. F. Tannaci, M. Noji, J. McBee, T. D. Tilley, *J. Org. Chem.* **2007**, *72*, 5567–5573.
[176] P. Solanke, F. Bureš, *Synthesis* **2013**, *45*, 3044–3051.

B7. As a first approach toward a 1,2-dihydroxy-3,6-diiodobenzene derivative suitable for the formation of a boronate ester, we purchased commercial 1,2-dichloro-4,5-dihydroxybenzene. Our plan was to protect the diol unit as an acetonide and then perform a iodination reaction at the two available aromatic positions (Scheme 8a). This route was explored as it was the most straightforward and the direct dehydration reaction between boronic acid and acetonide-protected catechol had previously been proven to work well.[177] Different protocols using both acetone and 2,2-dimethoxypropane in the presence of *p*-toluenesulfonic acid in a Dean-Stark apparatus did not work on 1,2-dichloro-4,5-dihydroxybenzene. Contrarily, the protection reaction of catechol as an acetonide was quantitative in the same conditions. The two chlorine atoms are believed to deactivate somehow the substrate. As the selective incorporation of only two iodine atoms *ortho* to the hydroxy groups could not be controlled, the following synthetic route toward 1,2-dihydroxy-3,6-diiodobenzene (**B7**) was considered (Scheme 8b).[178]



Scheme 8. (**a**) Attempted route toward 5,6-dichloro-4,7-diiodo-2,2-dimethyl-1,3-benzodioxole and (**b**) effective synthetic route to 1,2-dihydroxy-3,6-diiodobenzene (**B7**).

In this approach, 1,2-dihydroxy-3,6-diiodobenzene (**B7**) was obtained in four steps from 1,2-dimethoxybenzene (veratrol) with good overall yields. Firstly, 1,2-dimethoxy-3,6-bis(trimethylsilyl)benzene was prepared from veratrol in two steps that involved *ortho*-lithiation. This compound was then treated with ICl in CH_2Cl_2 to afford 1,4-diiodo-2,3-dimethoxybenzene. The two methyl groups were then removed in presence of the demethylating agent BBr_3.

1.2.2. Nucleobase Directors.

A series of chemically-modified lipophilic natural and non-natural nucleobase derivatives (Figure 51) that are equipped with either a halogen atom or an ethynyl group at

[177] E. L. Spitler, W. R. Dichtel, *Nat. Chem.* **2010**, *2*, 672–677.
[178] Y. L. Zhao, L. Liu, W. Zhang, C. H. Sue, Q. Li, O. Š. Miljanić, O. M. Yaghi, J. F. Stoddart, *Chem. Eur. J.* **2009**, *15*, 13356–13380.

the 5-position (for the pyrimidines) or at the 8-position (in the case of the purines) have been synthesized.[179]

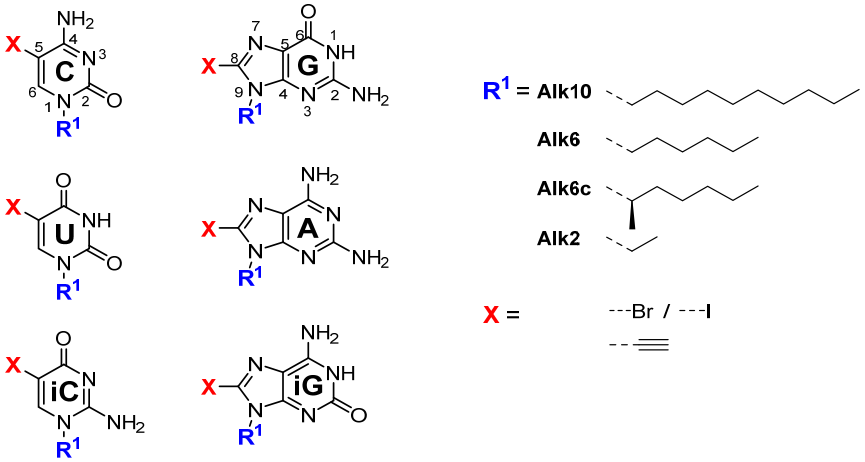


Figure 51. General structure of the targeted pyrimidine (C, U, iC) and purine (G, A, iG) derivatives.

As shown in Figure 51, the targeted nucleobases present common functional groups, such as the halogen atom or the ethynyl group, as well as variable substituents, such as the diverse lipophilic groups at the 1-position of pyrimidines and the 9-position of purines. Therefore, the ideal, most convergent synthetic route to these derivatives would comprise three consecutive steps: (1) halogenation (bromination or iodination); (2) Palladium-catalyzed Sonogashira coupling[169] with trimethylsilylacetylene (TMSA), followed by trimethylsilyl (TMS) deprotection; and (3) nucleophilic substitution with the corresponding alkylating reagent.[180] Since the variability in the compounds is introduced at the level of the alkyl substituents, a lower number of products would need to be synthesized using this sequence. This synthetic pathway was preferred and was always prioritized. However, we deviated from this optimal route whenever it was required because of different synthetic problems. Many intermediates were unreactive in some conditions and/or displayed very low solubility to be properly purified. The sequences and reactions described below in this Chapter have been optimized in order to reach the final compounds in the most straightforward and convenient way, and with the maximum purity and overall yields.

Synthesis of pyrimidines.

With the aim of reaching the most convergent route, iodination of the pyrimidine heterocycles was selected as the first reaction in the synthetic schemes toward lipophilic C, U and iC derivatives. The incorporation of iodine was chosen instead of bromine due to the

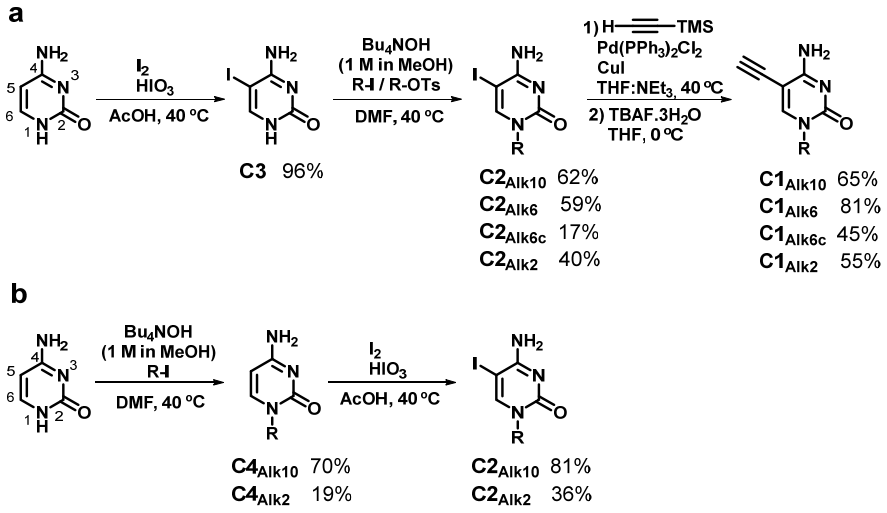
[179] N. Bilbao, V. Vázquez-González, M. Aranda, D. González-Rodríguez, *Eur. J. Org. Chem.* **2015**, *32*, 7160–7175.
[180] a) A. Holy, J. Gunter, H. Dvorakova, M. Masojidkova, G. Andrei, R. Snoeck, J. Balzarini, E. De Clercq, *J. Med. Chem.* **1999**, *42*, 2064–2086; b) M. N. S. Rad, A. Khalafi-Nezhad, S. Behrouz, M. A. Faghihi, A. Zare, A. Parhami, *Tetrahedron* **2008**, *64*, 1778–1785.

higher reactivity of iodoarenes in metal-catalyzed cross-coupling reactions. Different iodination agents and protocols, such as I_2/HIO_3, I_2/H_5IO_6, ICl or NIS, were essayed and the best option for each substrate was chosen. Next, the Sonogashira coupling in different conditions was attempted in each 5-iodopyrimidine. However, none of them (C, U or iC) was reactive, probably because *N*-1 unsubstituted pyrimidines can deactivate the catalytic palladium species through chelation *via* one of their different tautomeric forms, as has been suggested for other nucleobase derivatives.[181] 5-iodo-substituted pyrimidines were instead subjected first to an alkylation reaction which is, in most cases, quite selective for the *N*-1 position. The Sonogashira reaction was performed on these alkylated iodopyrimidines, leading to the coupled products. Finally, the TMS protecting group was cleaved in the presence of tetrabutylammonium fluoride (TBAF) or K_2CO_3.

Cytosine (Scheme 9). A total of eight 5-iodocytosine and 5-ethynylcytosine derivatives were synthesized, each of them having different lipophilic alkyl groups with variable length or presenting chiral centers. Iodination of commercial cytosine (Scheme 9a) was carried out in the presence of I_2 and HIO_3 and could be easily scaled up.[182] In these conditions, 5-iodocytosine (**C3**) was obtained in close to quantitative yields by neutralization and straightforward filtration and washing. This product was then subjected to an alkylation reaction in the presence of the corresponding alkyl halide or activated alcohol. These reactions are typically carried out in the presence of cesium or potassium carbonates but, in the case of the iodoalkanes, the use of these bases did not afford good yields. Instead, the more soluble Bu_4NOH base in methanol[183] was used to deprotonate **C3** and produce $\mathbf{C2_{Alk10}}$, $\mathbf{C2_{Alk6}}$, $\mathbf{C2_{Alk6c}}$ and $\mathbf{C2_{Alk2}}$ in acceptable yields. In order to reach $\mathbf{C2_{Alk6c}}$, (*S*)-(+)-2-heptanol, transformed into its tosylated form, was used. It is noteworthy that this substitution reaction mechanism takes place without epimerization of the steric center, since no planar carbocation intermediate is formed. Indeed, this alkylation reaction follows a S_N2 pathway resulting in the inversion of the chiral center.[184] Unfortunately, large amounts of the deiodinated product were collected as a byproduct in the reaction with these particular substrates. In addition to that, $\mathbf{C2_{Alk6c}}$ has a similar solubility than tetrabutylammonium salts and could not be effectively purified. However these impurities proved not to be an obstacle to reach $\mathbf{C1_{Alk6c}}$ through Sonogashira cross-coupling. 1-Alkyl-5-iodocytosines were transformed into the final 5-ethynyl derivatives after alkylation, using standard Sonogashira-TMS deprotection methods. Again $\mathbf{C1_{Alk6c}}$ could not be separated from tetrabutylammonium impurities because of its high solubility but did not hamper the

[181] a) E. C. Western, K. H. Shaughnessy, *J. Org. Chem.*, **2005**, *70*, 6378–6388; b) A. Collier, G. K. Wagner, *Org. Biomol. Chem.* **2006**, *4*, 4526–4532; c) A. Collier, G. K. Wagner, *Synth. Commun.* **2006**, *36*, 3713–3721.
[182] M. Bobek, I. Kavai, R. A. Sharma, S. Grill, G. Dutschman, Y.-C. Cheng, *J. Med. Chem.* **1987**, *30*, 2154–2157.
[183] E. Greco, A. E. Aliev, V. G. H. Lafitte, K. Bala, D. Duncan, L. Pilon, P. Golding, H. C. Hailes, *New J. Chem.* **2010**, *34*, 2634–2642.
[184] a) D. Catalano, L. Chiezzi, V. Domenici, M. Geppi, C. A. Veracini, *J. Phys. Chem. B* **2003**, *107*, 10104–10113; b) J. Elhaïk, C. M. Pask, C. A. Kilner, M. A. Halcrow, *Tetrahedron* **2007**, *63*, 291–298.

progress of the synthetic route. Final purification was carried out further on in the synthetic path (see section **2.4.2.** in Chapter 2).



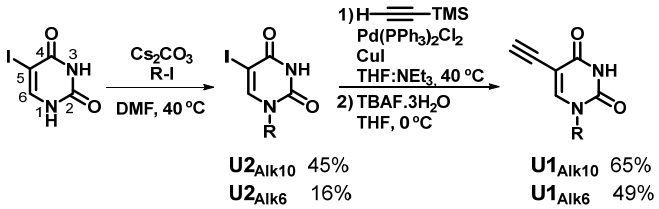
Scheme 9. Synthetic routes to 5-iodo- and 5-ethynyl-cytosines using the (**a**) halogenation-alkylation or (**b**) alkylation-halogenation sequence.

We also tested the inverted reaction order, that is alkylation of cytosine followed by halogenation (Scheme 9b). Besides being less convergent, this alternative route presented other types of drawbacks. First, the isolation of the halogenated products was more tedious and produced lower yields. Second, the presence of the bulky iodine atom at the C-5 position in **C3** was proven effective in directing alkylation selectively at the *N*-1 position.[185] For instance, when carrying out the alkylation reaction with ethyl iodide directly onto commercial cytosine, a triply alkylated byproduct was also obtained, having ethyl chains at the *N*-1 position but also in the amino group.

Uridine (Scheme 10). The route to the uridine derivatives started directly from commercial 5-iodouridine, a compound that can also be obtained in excellent yields from uridine.[186] 5-iodouridine was alkylated in the presence of Cs_2CO_3, which was proven to be slightly more efficient than K_2CO_3 because of its higher solubility in organic solvents. The yields of **U2$_{Alk10}$** and **U2$_{Alk6}$** were not very satisfactory, although they are in the same range of those reported on similar alkylation reactions with uridine derivatives.[187] One of the reasons for these relatively low yields is the additional formation of *N*-3-substituted products and the

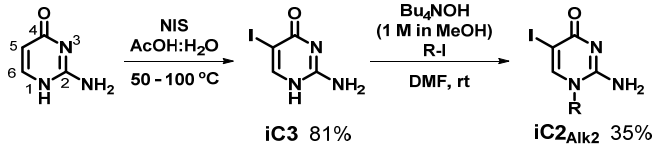
[185] D. L. Helfer II, R. S. Hosmane, N. J. Leonard, *J. Org. Chem.* **1981**, *46*, 4803–4804.
[186] R. H. E. Hudson, G. Li, J. Tse, *Tetrahedron Lett.* **2002**, *43*, 1381–1386.
[187] a) F. H. Beijer, R. P. Sijbesma, J. A. J. M. Vekemans, E. W. Meijer, H. Kooijman, A. L. Spek, *J. Org. Chem.* **1996**, *61*, 6371–6380; b) E. Coutouli-Argyropoulou, M. Tsitabani, G. Petrantonakis, A. Terzis, C. Raptopoulou, *Org. Biomol. Chem.* **2003**, *1*, 1382–1388; c) S. Prachayasittikul, N. Sornsongkhram, R. Pingaew, A. Worachartcheewan, S. Ruchirawat, V. Prachayasittikul, *Molecules* **2009**, *14*, 2768–2779; d) S. A. Shelke, G. B. Sandholt, S. T. Sigurdsson, *Org. Biomol. Chem.* **2014**, *12*, 7366–7374.

corresponding dialkylated products, identified by ^{1}H NMR and MS, which compete with alkylation at *N*-1. Finally, 1-alkyl-5-iodouracils were converted to the corresponding 5-ethynyl derivatives *via* cross-coupling reaction with TMSA. Traces of deiodinated products were detected in this last reaction.



Scheme 10. Optimized synthetic route to 5-iodo- and 5-ethynyl-uracils.

***Iso*cytosine** (Scheme 11). The unnatural *iso*cytosine analog **iC$_{Alk2}$** was also prepared following the halogenation-alkylation sequence. In this case, iodination was carried out in the presence of NIS,[188] since it afforded better results than other methods. Alkylation was then performed using the same procedure as with the cytosine derivatives. This reaction led as well to the *N*-3-substituted product in minor amounts, which could be separated by chromatography from the target *N*-1-alkylated derivative. Product **iC$_{Alk2}$** was not transformed into the corresponding 5-ethynyl derivative, as this nucleobase was intended for the further preparation of a short monomer, where the self-assembling units would be solely separated by an ethynyl group (see Chapter 3).



Scheme 11. Synthetic route to 5-iodo-isocytosine **iC2$_{Alk2}$**.

Synthesis of purines.

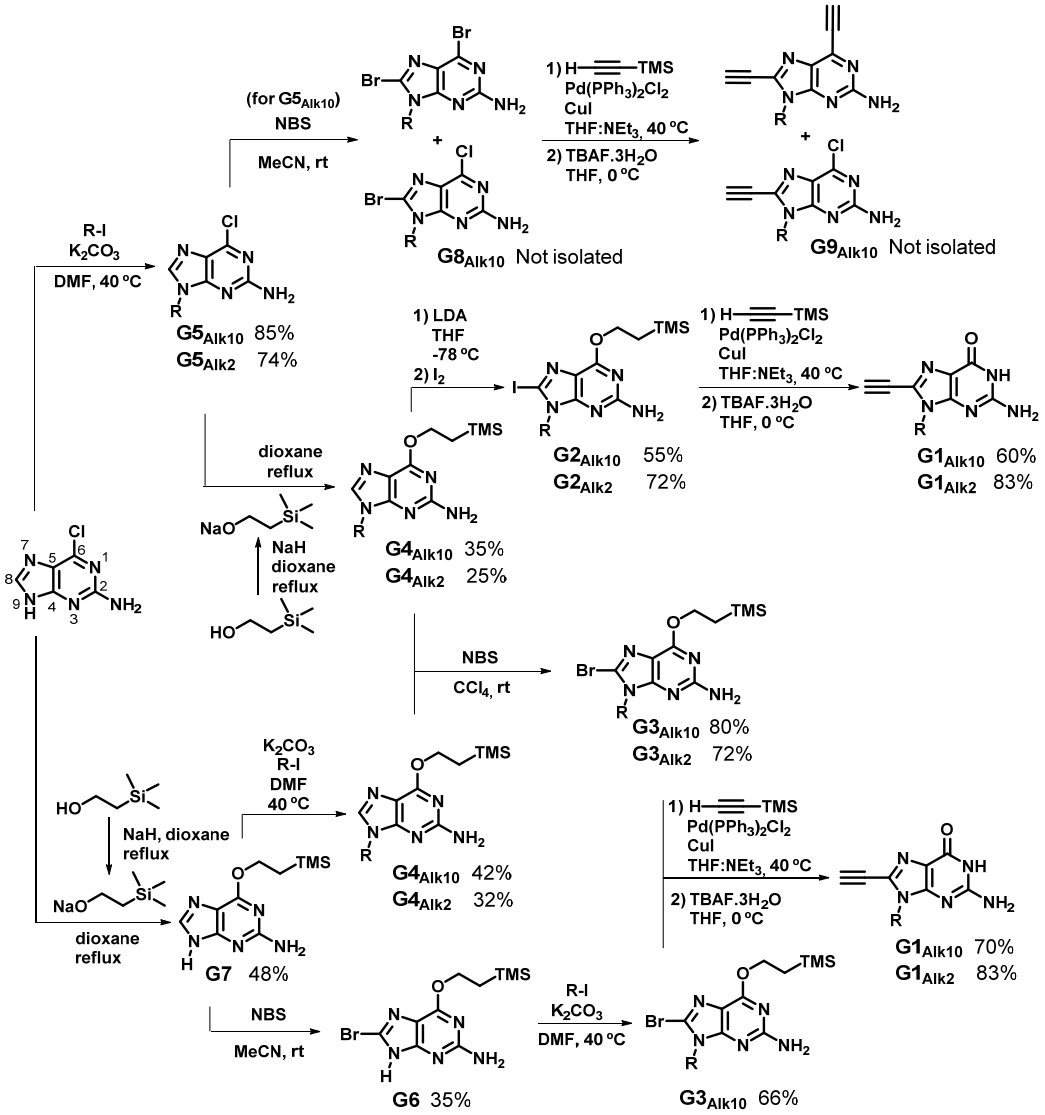
The optimized route to 8-halogenated and 8-ethynylated purines followed a different synthetic sequence to the one used with the pyrimidines. The most convergent path was again pursued and the halogenation step[189] wanted to be introduced as soon as possible in the sequence. Although this was accomplished in some cases, the early introduction of the halogen atom resulted in a series of secondary problems during subsequent reactions that reduced the overall yields. Hence, the alkylation reaction was preferably carried out as the first step in the routes to guanine, 2-aminoadenine and *iso*guanine derivatives, despite the higher number of products that needed to be synthesized. Then, the alkylated products

[188] A. Mayer, A. Häberli, C. J. Leumann, *Org. Biomol. Chem.* **2005**, *3*, 1653–1658.
[189] a) R. Volpini, D. D. Ben, C. Lambertucci, G. Marucci, R. C. Mishra, A. T. Ramadori, K.-N. Klotz, M. L. Trincavelli, C. Martini, G. Cristalli, *ChemMedChem* **2009**, *4*, 1010–1019; b) W. K.-D. Brill, C. Riva-Toniolo, *Tetrahedron Lett.* **2001**, *42*, 6279–6282.

were either brominated or iodinated. The preparation of 8-iodopurines was preferred because of their higher reactivity in subsequent Pd-catalyzed couplings, so their synthesis was optimized and studied in more detail using different iodination methods. The iodine substituent could be introduced either by electrophilic aromatic substitution in the presence of diverse reagents or by deprotonation at the 8-position followed by quenching with iodine. The preparation of 8-bromopurines was however also considered as an alternative. As with the pyrimidines, the Sonogashira reaction was left as the last step in the sequence to 8-ethynylpurines. This reaction, as will be explained below, required in some instances previous functional group protection of the purine heterocycles. As was the case with the pyrimidine bases, all our attempts to perform this palladium-catalyzed coupling after halogenation and before alkylation were unsuccessful, probably because of deactivation of the catalytic species due to a higher metal-coordination ability of *N*-9 unsubstituted derivatives.

Guanine (Scheme 12). The development of an efficient synthetic route toward 8-ethynyl lipophilic guanines (**G1$_{Alk10}$** and **G1$_{Alk2}$**) turned out to be a complex task and considerable effort was dedicated to this mission. The low solubility of guanine or 2-amino-6-hydroxypurine in different solvents is one of the several reasons. Among all the nucleobases, guanine exhibits the richest supramolecular chemistry,[55,56,127,190] and often self-associates in solution to form highly insoluble products or viscous gels. Also, guanine and its analogs are constituted by a mixture of tautomers that can lead to different isomers in the alkylation reactions. The most common of them are the *N*-7 and *N*-9 alkylated products, which are sometimes difficult to separate.[191] But probably the most important reason is the requirement to protect or mask the carbonyl group before Pd-catalyzed cross-couplings. As a matter of fact, the low reactivity of the guanine heterocyle in metal-mediated oxidative addition processes has been reported.[148,181,192] The low oxidation potential of this base or its ability to coordinate organometallic catalytic species are cited among the causes that would explain such lack of reactivity.[181] Conversion to different functionalized derivatives like 6-alkoxy[148,192,193] or 6-halogenopurines, which, at the same time, may exhibit a higher selectivity in the alkylation reaction, is an alternative. Unfortunately, the use of these starting reagents involve higher costs and an increase in the number of reaction steps, which results in poorer overall yields. Furthermore, previous methods reported in the literature for the conversion of 2-amino-6-chloropurine into guanine use vigorous acidic or basic conditions that are not compatible with sensitive functional groups.

[190] a) L. Ma, M. Iezzi, M. S. Kaucher, Y.-F. Lam, J. T. Davis, *J. Am. Chem. Soc.* **2006**, *128*, 15269–15277; b) J. E. Betancourt, M. Martín-Hidalgo, V. Gubala, J. M. Rivera, *J. Am. Chem. Soc.* **2009**, *131*, 3186–3188; c) M. C. Rivera-Sánchez, I. Andújar-de-Sanctis, M. García-Arriaga, V. Gubala, G. Hobley, J. M. Rivera, *J. Am. Chem. Soc.* **2009**, *131*, 10403–10405.
[191] R. Zou, M. J. Robins, *Can. J. Chem.* **1987**, *65*, 1436–1437.
[192] W. Elizabeth, R. D. Jonathan, M. J. Edward, M. G. Peter, H. S. Kevin, *J. Org. Chem.* **2003**, *68*, 6767–6774.
[193] A. Dumas, N. W. Luedtke, *Chem. Eur. J.* **2012**, *18*, 245–254.



Scheme 12. Synthetic routes to 8-iodo- and 8-ethynyl-guanines from 2-amino-6-chloropurine.

The use of 2-amino-6-chloropurine as the starting reagent was considered and different routes and sequences from this commercial derivative were evaluated (Scheme 12). The initial idea was to perform the whole three-step reaction sequence directly on this kind of derivatives (Scheme 12, upper part). Since all the attempts to halogenate this substrate under different conditions were unsuccessful, the alkylation of 2-amino-6-chloropurine with diverse iodoalkanes was introduced as the first step of the route, to yield **G5$_{Alk10}$** and **G5$_{Alk2}$** with 1-iododecane or 1-iodoethane, respectively. Mixtures of *N*-7 and *N*-9 substituted products were obtained in an approximate 1:4 ratio. This is presumably due

to the low steric hindrance of the iodoalkanes that were used. However the overall yields were still very satisfactory.

Next, NBS-mediated bromination on **G5$_{Alk10}$** was tested, which unexpectedly afforded a mixture of brominated products, mainly including the target 2-amino-8-bromo-6-chloropurine and 2-amino-6,8-dibromopurine, in which the 6-chloro atom was replaced by bromine. Unfortunately, this mixture could not be separated. Hence, the mixture was subjected to a Sonogashira coupling to see if one of these 6- or 8-positions reacted selectively or if the products could be separated at the last step by chromatography. Despite all the attempts under different conditions, both positions seemed to be equally reactive and our target compound could not be efficiently separated from the mixture of ethynylated products. Furthermore, the formation of hardly separable mixtures led to poor overall yields and time-consuming purification processes in the guanine synthetic route. On the other hand, the hydrolysis of the mixture of brominated products in acidic conditions[194] led to mixtures of 8- and 6-oxopurines.

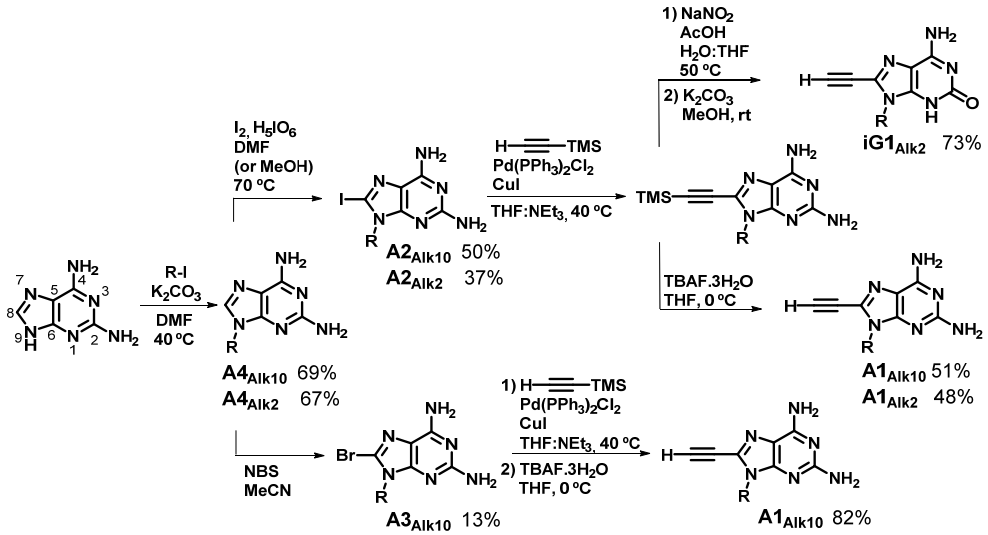
To circumvent these problems, substitution of the 6-chloro atom with a 2-trimethylsilylethoxy group[195] was next considered with the aim of reducing the reactivity of such position in the halogenation step (Scheme 12, lower part). This particular protecting group was chosen because of several additional advantages in the synthetic route: it imparts solubility to the heterocyclic ring and can be efficiently deprotected in the presence of fluoride at the very last step in the route,[196] along with the TMS group. Compound **G7**[196] was thus synthesized and subjected then to a halogenation reaction. Electrophilic iodination was first tested with negative results and the starting material was recovered in all cases. Bromination in the presence of Br_2 did not work properly either. Contrarily, bromination with NBS led to **G6**, which could be subsequently alkylated. It is worth to note that the alkylation reaction on this substrate led to complete regioselectivity at *N*-9, affording **G3$_{Alk10}$**. However, despite its convergence, this route was still not very satisfactory since the first two reactions (nucleophilic aromatic substitution and bromination) proceeded with low to moderate yields, were not very reproducible and the intermediate products obtained (**G7** and **G6**) could not be easily isolated from traces of impurities. Besides, **G6** was unreactive in Pd-catalyzed couplings, either in Sonogashira reactions with TMSA or in Stille reactions in the presence of ethynyltributylstannane. Altering the sequence by alkylation of **G7** led to more easily isolable products (**G4$_{Alk10}$** and **G4$_{Alk2}$**). The use of 1-iododecane or 1-iodoethane in this reaction led now to a *ca.* 1:1 mixture of *N*-7 and *N*-9 products, which is surprising in view of the higher regioselectivities attained in the alkylation reactions on **G6** or on 2-amino-6-chloropurine (see above). Bromination of **G4$_{Alk10}$** or **G4$_{Alk2}$** led to the 8-bromopurines **G3$_{Alk10}$** or **G3$_{Alk2}$** in quite good yields.

[194] J. Thibon, L. Latxague, G. Deleris, *J. Org. Chem.* **1997**, *62*, 4635–4642.
[195] F.-P. Sun, T. Darbre, *Helv. Chim. Acta* **2002**, *85*, 3002–1018.
[196] L. L. Gundersen, T. Benneche, F. Rise, A. Gogoll, K. Undheim, *Acta Chem. Scand.* **1992**, *46*, 761–771.

Nevertheless, the best route found to the target lipophilic guanines is shown in the central region of Scheme 12. Alkylation was first performed on 2-amino-6-chloropurine using K_2CO_3 or Cs_2CO_3 as the base. This reaction afforded rather soluble products with good yields that could be easily purified and subjected to the aromatic nucleophilic substitution with the 2-trimethylsilylethoxy group, leading to compounds **G4$_{Alk10}$** and **G4$_{Alk2}$**. The **G4$_{Alk10}$** and **G4$_{Alk2}$** products could be brominated at the 8-position in the presence of NBS. Also, iodination was essayed under a different mechanism that involved deprotonation at the guanine 8-position in the presence of LDA and quenching of the resulting anion with iodine.[197] Products **G2$_{Alk10}$** and **G2$_{Alk2}$** were in this way generated with acceptable yields. Finally, bromo-derivatives **G3$_{Alk10}$** and **G3$_{Alk2}$** and iodoguanines **G2$_{Alk10}$** and **G2$_{Alk2}$** were subjected to a Sonogashira coupling in the usual conditions, followed by deprotection of the alkyne TMS group, to yield **G1$_{Alk10}$** and **G1$_{Alk2}$**.

2-Aminoadenine (Scheme 13). Two alkylated adenine derivatives were prepared from commercial 2,6-diaminopurine. The alkylation reaction[198] was carried out as the first step of the synthetic route. Different bases can be used, such as NaH, Cs_2CO_3 or K_2CO_3. The latter (K_2CO_3) proved to be the best base for this reaction due to handling conditions, price and yield reasons. Compounds **A4$_{Alk10}$** and **A4$_{Alk2}$** were prepared in this way. Next, the halogenation reaction was optimized. NBS was attempted first and **A4$_{Alk10}$** afforded the brominated product **A3$_{Alk10}$** with a poor yield, whereas the bromination of **A4$_{Alk2}$** could not be achieved with NBS. Therefore, iodination in the presence of periodic acid[174] was chosen over the bromination reaction with NBS because of the poor yields and low reliability of the latter reaction, even with recrystallized NBS. It is worth to mention that none of the guanine derivatives tested could be iodinated under these conditions. Compounds **A2$_{Alk10}$**, **A2$_{Alk2}$** and **A3$_{Alk10}$** were then subjected to the Sonogashira-TMS cleavage protocol, leading to **A1$_{Alk10}$** and **A1$_{Alk2}$** in good yields. These products, especially the ethyl-aminoadenine derivatives, displayed the poorest solubility in most solvents compared to other purines. This is probably due to strong self-associating phenomena, as these compounds present five available H-bonding sites.

[197] P. Lang, G. Magnin, G. Mathis, A. Burger, J.-F. Biellmann, *J. Org. Chem.* **2000**, *65*, 7825–7832.
[198] L. Zhang, J. Fan, K. Vu, K. Hong, J.-Y. Le Brazidec, J. Shi, M. Biamonte, D. J. Busch, R. E. Lough, R. Grecko, Y. Ran, J. L. Sensintaffar, A. Kamal, K. Lundgren, F. J. Burrows, R. Mansfield, G. A. Timony, E. H. Ulm, S. R. Kasibhatla, M. F. Boehm, *J. Med. Chem.* **2006**, *49*, 5352–5362.



Scheme 13. Optimized synthetic route to 8-iodo- and 8-ethynyl-diaminopurine and 8-ethynyl-*iso*guanine.

Isoguanine (Scheme 13). The *iso*guanine non-natural nucleobase could be obtained from 2,6-diaminopurine in a single step by selective hydrolysis at C-2 *via* a diazonium intermediate.[199] For the sake of simplicity and convergence, we performed this process after the cross-coupling reaction of the iodinated **A2** derivatives and before TMS deprotection, since the terminal ethynyl group was found to interfere in this reaction. Compound **iG1$_{Alk2}$** was isolated in 73% overall yield from **A2$_{Alk2}$** after TMS-cleavage in the presence of K_2CO_3.

[199] a) S. C. Jurczyk, J. T. Kodra, J.-H. Park, S. A. Benner, T. R. Battersby, *Helv. Chim. Acta* **1999**, *82*, 1005–1015; b) X. Shi, J. C. Fettinger, M. Cai, J. T. Davis, *Angew. Chem. Int. Ed.* **2000**, *39*, 3124–3127.

1.3. Conclusions.

Four π-conjugated molecules presenting a large flat aromatic area disubstituted in opposite directions with halogen atoms (-I, -Br and –Cl) were synthesized to complete a useful collection of seven central blocks for further purposes within this Thesis. Building blocks **B1**-**B5** are well-suited to afford stability on planar substrates. They present interesting electronic properties, especially the anthracene and octafluoroanthracene central blocks, useful for further microscopy studies. Building blocks **B6** and **B7** present anhydride and 1,2-diol moieties, respectively. Both of these functionalities are reactive toward condensation *via* dehydration, in the presence of boronic acids to yield boronic esters; or in the presence of amines to yield imides, respectively. These starting products will serve us in Chapter 5 for the preparation of covalently fused-monomers.

On the other hand, a series of lipophilic nucleobases have been prepared, comprising natural and non-natural derivatives, which are substituted at the 5- (pyrimidines) or 8-position (purines) with either a halogen atom or a terminal triple bond. These include cytosine, *iso*cytosine and uracil as pyrimidine heterocycles, and guanine, *iso*guanine, and 2-aminoadenine as complementary purine bases. The *N*-1/*N*-9 position of these compounds was functionalized with different alkyl groups. This kind of functional groups are suited to afford solubility and direct nucleobase assembly in organic solvents or onto surfaces, as it will be further explained in this Thesis. The synthetic sequences leading to the final ethynylated compounds have been optimized for each base attending to convergence, convenience, ease of purification and overall yields. The results indicate that the choice of reaction conditions and the order in which the halogenation, alkylation and Sonogashira steps are performed is not always trivial and each base requires a particular optimized protocol.

The molecules prepared can be regarded as a relevant collection of “supramolecular synthons” and they will be useful for the purposes of this Thesis. Also, since a wide diversity of functional units may be attached by means of metal-catalyzed cross-coupling or “click” reactions with the ethynyls groups,[200] other uses can be contemplated in order to guide their organization. Besides, nucleobase derivatives play a crucial role in nature and have been found as clinically useful molecules with diverse biological activities. An example of interest is constituted by *N*-9 substituted guanines, such as *acyclovir*, because of their potential antiviral activity.[201] Also the cytidine analog *lamivudine* has been proven useful for the treatment of chronic hepatitis B and HIV/AIDS.[202] Thymine analog *zidovudine* is another type of antiretroviral drug used for the treatment of HIV/AIDS,[203] as the adenosine

[200] *Click Chemistry in Glycoscience: New Developments and Strategies* (Eds.: Z. J. Witczak, R. Bielski), John Wiley & Sons, New York, **2013**.
[201] R. Snoeck, *Int. J. Antimicrob. Agents* **2000**, *16*, 157–159.
[202] M. J. Koziel, M. G. Peters, *N. Engl. J. Med.* **2007**, *356*, 1445–1454.
[203] K. Wright, *Nature* **1986**, *323*, 283.

analog *abacavir* is.[204] For these reasons, the development of efficient and reliable synthetic routes to modified nucleobases as versatile building blocks in self-assembly can be of great importance and utility to pharmaceutical and supramolecular chemists.

[204] S. Mallal, D. Nolan, C. Witt, G. Masel, A. M. Martin, C. Moore, D. Sayer, A. Castley, C. Mamotte, D. Maxwell, I. James, F. T. Christiansen, *Lancet* **2002**, *359*, 727–732.

1.4. Experimental Section.

1.4.1. General Methods.

Chemicals were purchased from commercial suppliers and used without further purification. Solid, hygroscopic reagents were dried in a vacuum oven before use. *N*-Bromosuccinimide (NBS) was recrystallized from water. Reaction solvents were thoroughly dried before use employing standard methods or using a solvent purification system Innovative Technology Inc. MD-4-PS (CH_3CN, THF, CH_2Cl_2, Et_2O).

Chromatography: reactions were monitored by *Thin Layer Chromatography* (TLC) using 0.2 mm aluminium sheets precoated with silica gel 60 F254 (Merck). TLC plates were inspected with a UV lamp featuring both long-wave UV light (365 nm) and short-wave UV-light (254 nm).
Column chromatography was carried out on silica gel *Merck-60* (230-400 mesh, 60 Å). Eluent relative volume/volume ratio are indicated in each case.

Nuclear Magnetic Resonance Spectroscopy (NMR)**:** ^{1}H NMR, ^{13}C NMR and ^{19}F NMR spectra were recorded with a *BRUKER AC-300* (300 MHz) instrument at the Department of Organic Chemistry or a *BRUKER XRD-500* (500 MHz) instrument in the interdepartmental service for research (SIdI) at the UAM. The temperature was actively controlled at 298 K. Chemical shifts (δ) are measured in parts-*per*-million (ppm) using the signals of the deuterated solvent as the internal standard [$CHCl_3$, calibrated at 7.26 ppm (^{1}H) and 77.0 ppm (^{13}C); DMSO calibrated at 2.50 ppm (^{1}H) and 39.5 ppm (^{13}C)]. The used deuterated solvents are indicated in each case. The following abbreviations have been used for the spectra description: s (singlet), d (doublet), t (triplet), q (quartet), quint (quintet), m (multiplet), dd (doublet of doublets), bs (broad signal).

Mass Spectrometry (MS) and **High Resolution-Mass Spectrometry** (HRMS) spectra were measured in the SIdI on a *VG AutoSpec* apparatus (FAB) or an *Applied Biosystems QSTAR* equipment (ESI) in the positive or negative modes. MALDI-TOF-MS spectra were obtained from a BRUKER ULTRAFLEX III instrument equipped with a nitrogen laser operating at 337 nm. Gas Chromatography spectra were obtained from CP-3800 instrument equipped with a Varian 1200 L Quadrupole MS detector

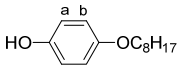
The **angle of rotation** ($[\alpha]_D^T$) of optically active substances were measured in a Perkin Elmer 241MC polarimeter at room temperature using a sodium D line with a light wavelength λ = 589 nm. The concentration (in g/mL) and solvent used are indicated in each case.

Melting points were determined in open-end capillary tubes in a Stuart melting point apparatus SMP3.

1.4.2. Synthesis and Characterization.

The synthesis and characterization of compounds 9,10-dichlorooctafluoroanthracene **B5**,[175] 5-iodocytosine **C3**[205] and 5-iodo-*iso*cytosine **iC3**[188] have been reported elsewhere. The NMR spectra of most of the nucleobase compounds described below can be found in the Supporting Information of our paper: *Eur. J. Org. Chem.* **2015**, *32*, 7160–7175 (ref.206).

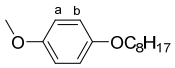
Synthesis of the dihalogenated central blocks.



4-octyloxyphenol. 4-octyloxyphenol was synthesized according to a literature procedure[174] adapted to our molecule. To a solution of hydroquinone (27.5 g, 250 mmol) and potassium hydroxide (21.0 g, 375 mmol) in dry DMF (300 mL), 1-bromooctane (8.7 mL, 50 mmol) was added dropwise at 0 °C. The mixture was stirred at room temperature overnight. Once completed, the mixture was cooled down to 0 °C and the excess of KOH was neutralized with HCl 1M. The mixture was then extracted with CH_2Cl_2, washed with water (3 x 200 mL) and the organic layer was dried over $MgSO_4$. After the solvent had been removed by rotary evaporation, the resulting crude was purified by column chromatography on silica gel eluted with hexane/ethyl acetate (5:1). 4-octyloxyphenol was obtained as a white solid (5.9 g, 53%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.76 (m, 4H, Ar), 4.42 (bs, 1H, O*H*), 3.89 (t, *J* = 6.5 Hz, 2H, $OCH_2C_7H_{15}$), 1.75 (m, 2H, $OCH_2CH_2C_6H_{13}$), 1.5-1.2 (m, 10H, $OC_2H_4C_5H_{10}CH_3$), 0.88 (t, *J* = 6.9 Hz, 3H, $OC_7H_{14}CH_3$).

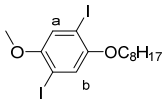
^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 153.4, 150.1, 116.2, 159.9, 69.3, 32.3, 23.1, 14.6.



1-methoxy-4-octyloxybenzene. Dry DMF (100 mL) was added over a system containing 4-octyloxyphenol (6.0 g, 27 mmol) and potassium carbonate (11.2 g, 81 mmol) under Ar. The mixture was stirred at room temperature for a period of 1 h. Then, methyl iodide (2.5 mL, 40.5 mmol) was added and the resulting mixture was stirred at room temperature overnight. Once completed, the solvent was removed by rotary evaporation, the crude was dissolved in $CHCl_3$ and washed with water (3 x 50 mL). The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure. The resulting crude was purified by column chromatography on silica gel eluted with hexane/ethyl acetate (20:1). 1-methoxy-4-octyloxybenzene was obtained as a white solid (5.2 g, 81%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 6.83 (s, 4H, Ar), 3.90 (t, *J* = 6.5 Hz, 2H, $OCH_2C_7H_{15}$), 3.77 (s, 3H, OCH_3), 1.75 (m, 2H, $OCH_2CH_2C_6H_{13}$), 1.5-1.2 (m, 10H, $OC_2H_4C_5H_{10}CH_3$), 0.89 (t, *J* = 6.9 Hz, 3H, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 157.9, 149.8, 113.0, 111.8, 69.0, 55.9, 31.9, 29.5, 29.34, 29.29, 26.1, 22.8, 16.1, 14.2.



1,4-diiodo-2-methoxy-5-octyloxybenzene (B2).[174] Iodine (2.3 g, 9.3 mmol) was added to a suspension of H_5IO_6 (1.4 g, 6.2 mmol) in MeOH (10 mL) at rt. After 10 min, a solution of 1-methoxy-4-octyloxybenzene (1.5 g, 6.2 mmol) in MeOH (10 mL) was added. The reaction mixture was stirred at 70 °C for 5 h. After completion, the reaction mixture was let to cool down to room temperature and $Na_2S_2O_5$ (sat) (10 mL) was added. The mixture was extracted in CH_2Cl_2 (30 mL) and washed with water (20 mL) and brine (20 mL). The organic layer was dried over

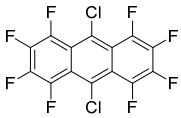
[205] L. M. Patil, D. E. Ponde, S. D. Samant, *Heterocycl. Lett.* **2014**, *4*, 559–564.
[206] http://dx.doi.org/10.1002/ejoc.201501026.

$MgSO_4$ and concentrated *in vacuo*. The crude solid was purified by column chromatography on silica gel eluted with hexane. **B2** was obtained as a yellow solid (2.6 g, 86%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.19 (s, 1H, H^a), 7.18 (s, 1H, H^b), 3.93 (t, J = 6.5 Hz, 2H, $OCH_2C_7H_{15}$), 3.82 (s, 3H, OCH_3), 1.80 (m, 2H, $OCH_2CH_2C_6H_{13}$), 1.50 (m, 2H, $OC_2H_4CH_2C_5H_{11}$), 1.4-1.2 (m, 8H, $OC_3H_6C_4H_8CH_3$), 0.89 (t, J = 6.9 Hz, 3H, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 153.3, 153.0, 123.0, 121.5, 86.4, 85.4, 76.7, 70.4, 57.2, 31.8, 29.23, 29.15, 26.0, 22.7, 14.1.

HRMS (GC-EI): Calculated for $C_{15}H_{23}O_2I_2$ $[M+H]^+$: 488.9788. Found: 488.9720.

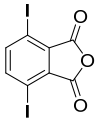


9,10-dichlorooctafluoroanthracene (B4). Compound **B4** was synthesized according to an adapted literature procedure.[175] A 25 mL flask was charged with octafluoroanthraquinone (0.5 g, 1.42 mmol) and phosphorus pentachloride (1.1 g, 5.11 mmol) and put under Ar. Phenylphosphonic dichloride (2 mL, 14.2 mmol) was added and the reaction was initially heated to 130 °C for 2 h followed by stirring at 170 °C for 24 h. The resulting orange solution was diluted with toluene and was successively washed with water (2 x 20 mL), $NaHCO_3$ (sat) (2 x 20 mL) and brine (20 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo* to a yellow oil. Without further purification, the crude oil was put in a flask equipped with a reflux condenser and NMP (11.6 mL, 120 mmol) and glacial AcOH (4 mL) were added. The solution was stirred under Ar at 115 °C for 24 h and the resulting dark-brown mixture was cooled to room temperature. The organic products were extracted with $CHCl_3$ and washed with water (2 x 10 mL), $NaHCO_3$ (sat) (2 x 10 mL) and brine (10 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo* to a brown oil. The crude oil was purified by column chromatography on silica gel eluted with hexane. **B4** was obtained as a yellow powder (0.15 g, 27%).

^{19}F NMR (300 MHz, $CDCl_3$) δ(ppm) = -137.09 (d, J = 12 Hz, 4F), -152.35 (d, J = 12 Hz, 4F).

^{13}C NMR (75 MHz, toluene-d_8) δ(ppm) = 144.9, 141.4, 132.9, 119.8.

MS (MALDI, dithranol): 393.4 $[M+H]^+$.

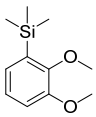


3,6-diiodophthalic anhydride (B6). B6 was synthesized according to an adapted literature procedure.[176] Into a solution of phthalic anhydride (10.0 g, 67.6 mmol) in fuming H_2SO_4 (50 mL, 120%) was added I_2 (34.4 g, 135.5 mmol) and the reaction mixture was stirred at 70 °C for 24 h. The mixture was poured over crushed ice (400 g), the precipitate was collected by filtration and washed with water (2 × 100 mL), $NaHCO_3$ (sat) (1 × 100 mL), $Na_2S_2O_3$ (sat) (1 × 100 mL), and washed again with water (1 × 100 mL). The crude product was extracted with CH_2Cl_2 (3 × 200 mL), the combined organic extracts were dried over $MgSO_4$ and evaporated. After recrystallization in acetic acid, compound **B6** was obtained as a yellow solid (4.06 g, 15%).

m. p. 229–233 °C.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) =7.92 (s, 2H, Ar).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 159.7, 147.5, 134.3, 91.3.

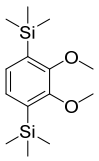


1,2-dimethoxy-3-trimethylsilylbenzene. 1,2-dimethoxy-3-trimethylsilylbenzene was synthesized according to an adapted literature procedure.[178,207] *n*BuLi (2.5 M, 43.5 mL, 108.8 mmol) was carefully added to a solution of veratrol (12.5 mL, 98.1 mmol) in dry THF (34 mL) and TMEDA (13.5 mL) at -78 °C. The mixture was stirred overnight at rt. Then, the mixture was cooled to -78 °C and TMSCl (15 mL, 118.2 mmol) was added dropwise. The resulting mixture was stirred at rt for 90 min and then quenched with water (40 mL). The water phase was separated and extracted with cyclohexane. The combined organic phases were dried over $MgSO_4$ and concentrated *in vacuo*. The crude

[207] D. W. Slocum, J. Ray, P. Shelton, *Tetrahedron Lett.* **2002**, *43*, 6071–6073.

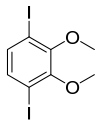
oil was purified by column chromatography on silica gel eluted with cyclohexane/CH_2Cl_2 (10:1), affording 1,2-dimethoxy-3-trimethylsilylbenzene as an uncolored oil (16.41 g, 80%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.12 – 6.93 (m, 3H, Ar), 3.88 (s, 3H, OCH_3), 3.87 (s, 3H, OCH_3), 0.30 (s, 9H, Si(CH_3)$_3$).



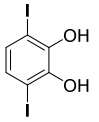
1,2-dimethoxy-3,6-bis(trimethylsilyl)benzene.[178] *n*BuLi (2.5 M, 38.5 mL, 96.25 mmol) was added dropwise to a solution of 1,2-dimethoxy-3-trimethylsilylbenzene (16.41 mL, 78.03 mmol) in TMEDA (14.3 mL) at 0 °C. The reaction mixture was stirred overnight at rt. Then, the resulting mixture was cooled down to -78 °C and TMSCl (14.5 mL, 114.25 mmol) was carefully added. The mixture was stirred at rt for 90 min. The reaction mixture was then neutralized with water (40 mL) and the water phase was separated and extracted with cyclohexane. The combined organic phases were dried over MgSO4 and concentrated *in vacuo*. The crude oil was purified by column chromatography on silica gel eluted with cyclohexane/CH_2Cl_2 (10:1), affording 1,2-dimethoxy-3,6-bis(trimethylsilyl)benzene as an uncolored oil (21.0 g, 95%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.11 (s, 2H, Ar), 3.85 (s, 6H, OCH_3), 0.28 (s, 18H, Si(CH_3)$_3$).



1,4-diiodo-2,3-dimethoxybenzene.[178] A solution of ICl (25.82 g, 159.0 mmol) in CH_2Cl_2 (120 mL) was added dropwise to a solution of 1,4-bis(trimethylsilyl)-2,3-dimethoxybenzene (21 g, 74.3 mmol) in CH_2Cl_2 (100 mL) at 0 °C. The mixture was stirred at rt for 30 min. Then, $Na_2S_2O_3$ (sat) (200 mL) was added to the mixture and the water phase was separated and extracted with CH_2Cl_2. The combined organic phases were concentrated under reduced pressure, dissolved in CH_2Cl_2 and washed with $Na_2S_2O_3$ (sat) and brine. The organic layer was then dried over MgSO4 and concentrated *in vacuo*, affording 1,4-diiodo-2,3-dimethoxybenzene as yellow solid (29.81 g, 99%).

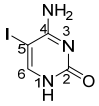
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.23 (s, 2H, Ar), 3.86 (s, 6H, OCH_3).



1,2-dihydroxy-3,6-diiodobenzene (B7).[178] A solution of BBr_3 in hexanes (1 M, 59 mL, 59 mmol) was added dropwise to a solution of 1,4-diiodo-2,3-dimethoxybenzene (5.08 g, 13.02 mmol) in CH_2Cl_2 (10 mL) at -78 °C. The reaction mixture was stirred overnight at rt. Then, the mixture was poured over ice (150 g) and the water phase was separated and extracted with ethyl acetate. The combined organic phases were dried over $MgSO_4$ and concentrated *in vacuo*. The crude oil was purified by column chromatography on silica gel eluted with $CHCl_3$, affording **B7** as a brown oil (3.98 g, 84%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) =7.00 (s, 2H, Ar), 5.66 (s, 2H, O*H*).

Synthesis of the cytosine derivatives.



5-iodocytosine (C3). C3 was synthesized according to a literature procedure[182] adapted to our molecule. Cytosine (10.0 g, 90.0 mmol), iodine (34.3 g, 135.0 mmol) and iodic acid (22.2 g, 126.0 mmol) were stirred in acetic acid (300 mL) at 40 °C overnight. Once completed, the reaction mixture was cooled and treated with $Na_2S_2O_3$ (sat) (200 mL) until a white suspension was obtained. The mixture was then neutralized with NaOH 6 M (900 mL). A white solid was formed that was filtered off and washed with slightly basified water until the filtered water had neutral pH. **C3** was dried under reduced pressure affording 20.6 g (96% yield).

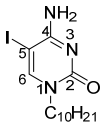
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.76 (bs, 2H, N^1*H*), 7.76 (s, 1H, *H*^{6}), 6.48 (bs, 2H, NH_2).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.4, 155.8, 149.4, 55.2.

HRMS (ESI+): Calculated for $C_4H_5N_3OI$ $[M+H]^+$: 237.9471. Found: 237.9464.

- ***Standard Procedure A*** for the nucleobase alkylation reaction. To a suspension of the nucleobase starting material (1 eq) and a base (1.2 eq) (indicated in each case) in dry DMF (volume indicated in each case) was added dropwise the corresponding iodoalkane (1.2 eq) (indicated in each case). The mixture was

stirred under argon at 40 °C for a period of time (indicated in each case) until completion, which was monitored by TLC. Work-up and purification methods are indicated in each case.



1-decyl-5-iodocytosine ($C2_{Alk10}$). $C2_{Alk10}$ can be synthesized according to a literature procedure[183] adapted to our molecule by following *Standard Procedure A* using **C3** (10.0 g, 42.2 mmol), a 1.0 M solution in MeOH of Bu_4NOH (50.6 mL, 50.6 mmol), 1-iododecane (11.0 mL, 50.6 mmol) and DMF (150 mL). The reaction was completed in 12 h. Then, the reaction mixture was poured into 150 mL of water and the precipitated solid was filtered, washed with water and dried. The resulting solid was washed with acetonitrile affording 9.8 g of $C2_{Alk10}$ (62% yield).

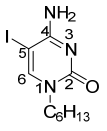
$C2_{Alk10}$ can also be obtained by stirring $C4_{Alk10}$ (15.7 g, 62.5 mmol), iodine (10.1 g, 40.5 mmol) and iodic acid (13.2 g, 76.0 mmol) in 400 mL acetic acid at 40 °C overnight. Once completed, the reaction mixture was cooled and the insoluble iodic acid was filtered away. The resulting mixture was extracted with ethyl acetate and washed with water (3 x 400 mL), $NaHCO_3$ (sat) (3 x 400 mL), $Na_2S_2O_3$ (sat) (400 mL) and brine (400 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The resulting solid was washed with ether affording $C2_{Alk10}$ as a bright beige solid (15.8 g, 81%).

m. p. 163–165 °C

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.07 (s, 1H, H^6), 7.58 (bs, 1H, NH-*H*), 6.43 (bs, 1H, NH-*H*), 3.7-3.6 (m, 2H, $N^1CH_2C_9H_{19}$), 1.6-1.4 (m, 2H, $N^1CH_2CH_2C_8H_{17}$), 1.4-1.1 (m, 14H, $N^1C_2H_4C_7H_{14}CH_3$), 0.85 (t, J = 7.0 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.8, 155.4, 150.9, 55.0, 50.5, 31.8, 29.5, 29.4, 29.3, 29.23, 29.17, 26.5, 22.6, 14.1.

HRMS (ESI+): Calculated for $C_{14}H_{25}N_3OI$ $[M+H]^+$: 378.1036. Found: 378.1028.



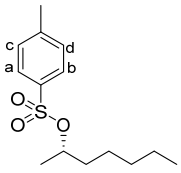
1-hexyl-5-iodocytosine ($C2_{Alk6}$). $C2_{Alk6}$ can be obtained following *Standard Procedure A* using **C3** (5.0 g, 21.1 mmol), a 1.0 M solution in MeOH of Bu_4NOH (25 mL, 25.3 mmol), 1-iodohexane (4 mL, 25.3 mmol) and DMF (75 mL). The reaction was completed in 8 h. The reaction mixture was then poured into 75 mL of water and the precipitated solid was filtered, washed with water and dried. The resulting solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording 4.0 g of $C2_{Alk6}$ (59% yield) as a pale solid.

m. p. 165–166 °C.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.39 (bs, 1H, NH-*H*), 7.54 (s, 1H, H^6), 5.54 (bs, 1H, NH-*H*), 3.73 (m, 2H, $N^1CH_2C_5H_{11}$), 1.8-1.6 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.4-1.2 (m, 6H, $N^1C_2H_4C_3H_6CH_3$), 0.87 (t, J = 6.6 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 163.9, 154.6, 151.9, 55.0, 48.6, 30.8, 28.6, 25.5, 21.9, 13.8.

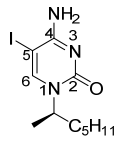
HRMS (ESI+): Calculated for $C_{10}H_{17}N_3OI$ $[M+H]^+$: 322.0410. Found: 322.0405.



(*S*)-2-heptanyl-4-methylbenzene sulfonate. A solution of (*S*)-(+)-2-heptanol (2.0 mL, 14 mmol) in dry CH_2Cl_2 (20 mL) was slowly added *via cannula* over a solution of 4-toluenesulfonyl chloride (4.00 g, 21 mmol) and NEt_3 (3.9 mL, 28 mmol) in dry CH_2Cl_2 (20 mL) at 0 °C. The reaction mixture was first stirred at 0 °C for 2 h and then at room temperature for 2 h more. Once completed, the mixture was washed with water (20 mL), with diluted HCl (aq) (20 mL), $NaHCO_3$ (sat) (20 mL) and brine (20 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo* to a yellowish oil. The crude oil was purified by column chromatography on silica gel eluted with hexane/ethyl acetate (100:1). (*S*)-2-heptanyl-4-methylbenzene sulfonate was obtained as a pale oil (0.38 g, 10%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.79 (d, J = 8.3 Hz, 2H, H^c, H^d), 7.33 (d, J = 8.3 Hz, 2H, H^a, H^b), 4.60 (sext, J = 6.2 Hz, 1H, C*H*), 2.44 (s, 3H, -Ar-CH_3), 1.54 (m, 2H, CH(OTs)$CH_2(C_4H_9)$), 1.26 (d, J = 7.0 Hz, 3H, CH(OTs)CH_3), 1.16 (m, 6H, $CH_2(C_3H_6)CH_3$), 0.82 (t, J = 7.0 Hz, 3H, $(C_4H_8)CH_3$).

$[\alpha]_D^{20}$ = + 13.0 (c = 6.9 x 10^{-4} g/mL, CH_2Cl_2).

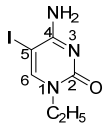


(2-(*R*)-methyl)-1-hexyl-5-iodocytosine ($C2_{Alk6c}$). $C2_{Alk6c}$ can be obtained following *Standard Procedure A* using **C3** (0.30 g, 1.28 mmol), a 1.0 M solution in MeOH of Bu_4NOH (1.5 mL, 1.54 mmol) and (*S*)-2-heptanyl-4-methylbenzene sulfonate (0.38 g, 1.41 mmol) in DMF (10 mL). The reaction was completed in 8 h. The reaction mixture was concentrated *in vacuo* and the resulting solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). A final recrystallization with cold acetonitrile afforded $C2_{Alk6c}$ as a white solid (0.71 g, 17% yield). This yield has to be considered as a maximum value, since $C2_{Alk6c}$ was contaminated with a small amount of tetrabutylammonium salts.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.52 (s, 1H, H^6), 6.82 (bs, 1H, NH-*H*), 5.40 (bs, 1H, NH-*H*), 4.82 (sext, *J* = 6.8 Hz, 1H, C*H*(C_5H_{11})), 1.59 (m, 3H, N^1CHCH_3), 1.4-1.2 (m, 8H, $N^1CH(C_4H_8)CH_3$), 0.87 (t, *J* = 6.8 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 163.1, 155.4, 147.3, 55.8, 52.3, 36.0, 31.4, 25.7, 22.4, 20.3, 13.9.

HRMS (ESI+): Calculated for $C_{11}H_{19}N_3OI$ $[M+H]^+$: 322.0410. Found: 336.0569.



1-ethyl-5-iodocytosine ($C2_{Alk2}$). $C2_{Alk2}$ can be obtained following *Standard Procedure A* using **C3** (10.0 g, 42.2 mmol), a 1.0 M solution in MeOH of Bu_4NOH (50.6 mL, 50.6 mmol) and 1-iododecane (11 mL, 50.6 mmol) in DMF (150 mL). The reaction was completed in 5 h. After solvent evaporation, the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1). A final recrystallization using acetonitrile yielded $C2_{Alk2}$ as a white solid (2.2 g, 40%).

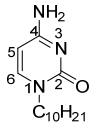
$C2_{Alk2}$ can also be obtained by stirring $C4_{Alk2}$ (1.0 g, 7.2 mmol), iodine (2.8 g, 11.1 mmol) and iodic acid (1.5 g, 8.8 mmol) in 20 mL acetic acid at 40 °C. The reaction was completed in 4 h. Then, the reaction mixture was cooled to room temperature and the insoluble iodic acid was filtered away. The mixture was extracted with $CHCl_3$ and washed with water (3 x 10 mL), $NaHCO_3$ (sat) (3 x 10 mL), $Na_2S_2O_3$ (sat) (10 mL) and brine (10 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The resulting solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1), affording $C2_{Alk2}$ as a white solid (0.7 g, 36%).

m. p. 203–204 °C.

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.09 (s, 1H, H^6), 7.61 (bs, 1H, NH-*H*), 6.46 (bs, 1H, NH-*H*), 3.67 (q, *J* = 7.0 Hz, 2H, $N^1CH_2CH_3$), 1.13 (t, *J* = 7.0 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 163.9, 154.4, 151.6, 55.2, 43.8, 14.5.

HRMS (ESI+): Calculated for $C_6H_9N_3OI$ $[M+H]^+$: 265.9784. Found: 265.9781.

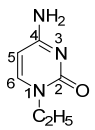


1-decylcytosine ($C4_{Alk10}$).[183] $C4_{Alk10}$ was obtained following *Standard Procedure A* using cytosine (10.0 g, 90 mmol), a 1.0 M solution in MeOH of Bu_4NOH (108 mL, 108 mmol), 1-iododecane (19 mL, 108 mmol) and DMF (500 mL). The reaction was completed in 12 h. Then, the reaction mixture was poured into 150 mL of water and the precipitated solid was filtered, washed with water and dried. The resulting solid was washed with acetonitrile affording $C4_{Alk10}$ as a pale solid (15.8 g, 70%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 7.54 (d, J = 7.0 Hz, 1H, H^6), 6.91 (bs, 2H, NH_2), 5.61 (d, *J* = 7.0 Hz, 1H, H^5), 3.59 (m, 2H, $N^1CH_2C_9H_{19}$), 1.53 (m, 2H, $N^1CH_2CH_2C_8H_{17}$), 1.3-1.1 (m, 14H, $N^1C_2H_4C_7H_{14}CH_3$), 0.85 (t, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 165.8, 155.8, 146.0, 93.0, 48.6, 38.5, 31.2, 28.9, 28.6, 25.9, 22.0, 13.9.

HRMS (ESI+): Calculated for $C_{14}H_{26}N_3O$ $[M+H]^+$: 252.2070. Found: 252.2082.



1-ethylcytosine ($C4_{Alk2}$). $C4_{Alk2}$ was obtained following *Standard Procedure A* using cytosine (5.0 g, 45 mmol), a 1.0 M solution in MeOH of Bu_4NOH (54 mL, 54 mmol), 1-iodoethane (4.5 mL, 54 mmol) and DMF (150 mL). The reaction was completed in 12 h. After solvent evaporation, the

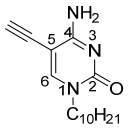
crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1). A final recrystallization using acetonitrile yielded **$C4_{Alk2}$** as a white solid (1.2 g, 19%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 7.59 (d, J = 7.0 Hz, 1H, H^6), 7.02 (bs, 2H, NH_2), 5.64 (d, J = 7.0 Hz, 1H, H^5), 3.65 (q, J = 7.0 Hz, 2H, N^{1}CH_2CH$_3$), 1.13 (t, J = 7.0 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 165.5, 155.2, 145.9, 93.3, 43.7, 14.5.

HRMS (ESI+): Calculated for $C_6H_{10}N_3O$ $[M+H]^+$: 140.0818. Found: 140.0813.

- ***Standard Procedure B*** for the Sonogashira coupling with TMSA and subsequent alkyne-TMS group deprotection. A dry THF/NEt_3 (4:1) solvent mixture was subjected to deoxygenation by three freeze-pump-thaw cycles with argon. Then, this solvent was added over the system containing the corresponding halogenated base (1 eq), CuI (0.01 eq) and $Pd(PPh_3)_2Cl_2$ (0.02 eq). The mixture was stirred at room temperature during a few minutes. Then, trimethylsilylacetylene (TMSA; 2 eq) was added dropwise. The reaction was stirred under argon at a given temperature for a period of time (indicated in each case) until completion, which was monitored by TLC. Then, the mixture was filtered over celite and the solvent evaporated under vacuum. The resulting crude was placed in a round-bottomed flask equipped with a magnetic stirrer, THF was added and the mixture was stirred at room temperature until the solid was dissolved. Then, hydrated tetrabutylammonium fluoride (TBAF·3H_2O; 1 eq) was added at 0 ^{o}C, and the mixture was stirred until reaction completion, which was monitored by TLC (approximately 1 hour in all cases). The solvent was evaporated at reduced pressure and the product was purified by column chromatography (eluent indicated in each case). The resulting solid was finally washed with cold acetonitrile.



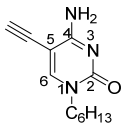
1-decyl-5-ethynylcytosine ($C1_{Alk10}$). $C1_{Alk10}$ was prepared following *Standard Procedure B*. **$C2_{Alk10}$** (10.0 g, 26.5 mmol), $Pd(PPh_3)_2Cl_2$ (372.1 mg, 0.53 mmol) and CuI (50.4 mg, 0.27 mmol) were dissolved in the THF/NEt_3 mixture (125 mL). Then TMSA (9.8 mL, 53.0 mmol) was added. The reaction was completed in 12 h. Then, TBAF·3H_2O (9.2 g, 29.2 mmol) was added over a THF (130 mL) solution of the crude mixture. **$C1_{Alk10}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/ethyl acetate (20:1). A final recrystallization using acetonitrile yielded **$C1_{Alk10}$** as a brown solid (4.7 g, 65%).

m. p. 139–141 ^{o}C.

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.05 (s, 1H, H^6), 7.50 (bs, 1H, NH-H), 6.66 (bs, 1H, NH-H), 4.30 (s, 1H, C≡CH), 3.7-3.6 (m, 2H, N^{1}CH_2C$_9$H$_{19}$), 1.6-1.4 (m, 2H, N^{1}CH$_2$CH_2C$_8$H$_{17}$), 1.3-1.1 (m, 14H, N^{1}C$_2$H$_4$C$_7$$H_{14}CH_3$), 0.85 (t, J = 6.6 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 164.9, 154.8, 148.8, 89.5, 83.7, 74.9, 50.6, 31.8, 29.4, 29.3, 29.2, 29.12, 29.10, 26.4, 22.6, 14.0.

HRMS (ESI+): Calculated for $C_{16}H_{26}N_3O$ $[M+H]^+$: 276.2070. Found: 276.2079.



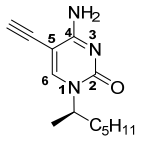
1-hexyl-5-ethynylcytosine ($C1_{Alk6}$). $C1_{Alk6}$ was prepared following *Standard Procedure B*. **$C2_{Alk6}$** (3.7 g, 11.6 mmol), $Pd(PPh_3)_2Cl_2$ (162 mg, 0.23 mmol) and CuI (22 mg, 0.12 mmol) were dissolved in the THF/NEt_3 mixture (70 mL). Then TMSA (3.3 mL, 23.1 mmol) was added. The reaction was completed in 12 h. Then, TBAF·3H_2O (3.7 g, 11.6 mmol) was added over a THF (60 mL) solution of the crude mixture. **$C1_{Alk6}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). A final recrystallization using acetonitrile yielded **$C1_{Alk6}$** as an ochre solid (2.1 g, 81%).

m. p. 163–164 ^{o}C.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.54 (s, 1H, H^6), 6.89 (bs, 1H, NH-H), 5.64 (bs, 1H, NH-H), 3.77 (m, 2H, N^{1}CH_2C$_5$H$_{11}$), 3.35 (s, 1H, C≡CH), 1.8-1.6 (m, 2H, N^{1}CH$_2$CH_2C$_4$H$_9$), 1.4-1.2 (m, 6H, N^{1}C$_2$H$_4$C$_3$$H_6CH_3$), 0.88 (t, J = 6.6 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.7, 154.8, 149.1, 89.2, 83.7, 75.0, 50.7, 31.4, 29.2, 26.2, 22.5, 14.0.

HRMS (FAB+): Calculated for $C_{12}H_{18}N_3O$ $[M+H]^+$: 220.1450. Found: 220.1445.

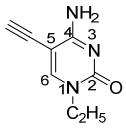


(2-(*R*)-methyl)-1-hexyl-5-ethynylcytosine ($C1_{Alk6c}$). $C1_{Alk6c}$ was prepared following *Standard Procedure B*. $C2_{Alk6c}$ (0.34 g, 1.03 mmol), $Pd(PPh_3)_2Cl_2$ (27.0 mg, 0.02 mmol) and CuI (2.0 mg, 0.01 mmol) were dissolved in the THF/NEt_3 mixture (10 mL). Then TMSA (0.3 mL, 2.06 mmol) was added. The reaction was completed in 3 h. Then, K_2CO_3 (0.21 g, 1.55 mmol) was added over a MeOH (10 mL) solution of the crude mixture. $C1_{Alk6c}$ was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). $C1_{Alk6c}$ was obtained as a dark solid unpurified by some tetrabuthylammonium salt residues (0.11 g, 45%). This yield has to be considered as a maximum value, since $C1_{Alk6c}$ was contaminated with a small amount of tetrabutylammonium salts.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.54 (s, 1H, H^6), 5.83 (bs, 1H, NH-*H*), 5.57 (bs, 1H, NH-*H*), 4.88 (q, *J* = 7.1 Hz, 1H, -C*H*(C_5H_{11})), 3.48 (s, 1H, C≡C*H*), 1.70 (m, 3H, N^1CHCH_3), 1.28 (m, 8H, $N^1CH(C_4H_8)CH_3$), 0.86 (t, *J* = 6.8 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 163.6, 154.3, 146.3, 131.7, 89.2, 79.5, 52.3, 35.7, 25.4, 22.2, 20.1, 13.7.

MS (ESI+): 465.2954 $[2M+H]^+$.



1-ethyl-5-ethynylcytosine ($C1_{Alk2}$). $C1_{Alk2}$ was prepared following *Standard Procedure B*. $C2_{Alk2}$ (1.5 g, 5.7 mmol), $Pd(PPh_3)_2Cl_2$ (79.5 mg, 0.11 mmol) and CuI (10.8 mg, 0.06 mmol) were dissolved in the THF/NEt_3 mixture (15 mL). Then TMSA (2 mL, 11.3 mmol) was added. The reaction was completed in 12 h. Then, TBAF·$3H_2O$ (1.8 g, 5.7 mmol) was added over a THF (50 mL) solution of the crude mixture. $C1_{Alk2}$ was purified by column chromatography on silica gel eluted with $CHCl_3$/ethyl acetate (20:1). A final recrystallization using acetonitrile yielded $C1_{Alk2}$ as a white solid (0.51 g, 55%).

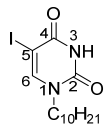
m. p. 196–197 °C

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.79 (bs, 1H, NH-*H*), 7.55 (s, 1H, H^6), 5.76 (bs, 1H, NH-*H*), 3.84 (q, *J* = 7.2 Hz, 2H, $N^1CH_2CH_3$), 3.35 (s, 1H, C≡C*H*), 1.32 (t, *J* = 7.2 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.7, 154.6, 148.6, 89.7, 83.8, 74.9, 45.6, 14.6.

HRMS (ESI+): Calculated for $C_8H_{10}N_3O$ $[M+H]^+$: 164.0818.Found: 164.0822.

Synthesis of the uracil derivatives.



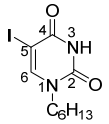
1-decyl-5-iodouracil ($U2_{Alk10}$). $U2_{Alk10}$ was obtained following *Standard Procedure A* using 5-iodouracil (5.0 g, 21 mmol), Cs_2CO_3 (8.2 g, 25.2 mmol), 1-iododecane (5.4 mL, 25.2 mmol) and DMF (100 mL). The reaction was completed in 4 h. After solvent evaporation, the crude was dissolved in $CHCl_3$ (200 mL) and washed with water (3 x 100 mL) and brine (100 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (200:1). $U2_{Alk10}$ was obtained as a white solid (3.6 g, 45%).

m. p. 148–149 °C

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.58 (bs, 1H, N^3H), 8.20 (s, 1H, H^6), 3.64 (t, *J* = 6.9 Hz, 2H, $N^1CH_2C_9H_{19}$), 1.55 (m, 2H, $N^1CH_2CH_2C_8H_{17}$), 1.3-1.1 (m, 14H, $N^1C_2H_4C_7H_{14}CH_3$), 0.85 (t, *J* = 6.9 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 160.8, 150.7, 150.0, 67.5, 49.1, 31.7, 29.3, 29.2, 29.1, 29.00, 28.96, 26.2, 22.5, 13.5.

HRMS (FAB+): Calculated for $C_{14}H_{24}N_2O_2I$ $[M+H]^+$: 379.0883. Found: 379.0881.

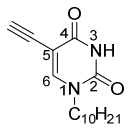


1-hexyl-5-iodouracil ($U2_{Alk6}$). $U2_{Alk6}$ was obtained following *Standard Procedure A* using 5-iodouracil (3.0 g, 12.6 mmol), Cs_2CO_3 (4.9 g, 15.1 mmol), 1-iodohexane (2.2 mL, 15.1 mmol) and DMF (200 mL). The reaction was completed in 4 h. After solvent evaporation, the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (200:1). $U2_{Alk6}$ was obtained as a white solid (0.64 g, 16%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.20 (bs, 1H, N^3H), 7.60 (s, 1H, H^6), 3.73 (t, J = 7.4 Hz, 2H, $N^1CH_2C_5H_{11}$), 1.7-1.6 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.3-1.2 (m, 6H, $N^1C_2H_4C_3H_6CH_3$), 0.90 (t, J = 7.0 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 169.5, 154.3, 148.9, 67.3, 49.2, 31.2, 29.1, 26.0, 22.4, 13.9.

HRMS (ESI+): Calculated for $C_{10}H_{16}N_2O_2I$ $[M+H]^+$: 323.0251. Found: 323.0249.

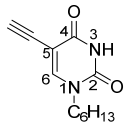


1-decyl-5-ethynyluracil ($U1_{Alk10}$). $U1_{Alk10}$ was prepared following *Standard Procedure B*. $U2_{Alk10}$ (3.2 g, 8.4 mmol), $Pd(PPh_3)_2Cl_2$ (117.6 mg, 0.17 mmol) and CuI (15.9 mg, 0.08 mmol) were dissolved in the THF/NEt_3 mixture (60 mL). Then TMSA (3.1 mL, 16.8 mmol) was added. The reaction was completed in 12 h. Then, TBAF·$3H_2O$ (2.7 g, 8.4 mmol) was added over a THF (60 mL) solution of the crude mixture. $U1_{Alk10}$ was purified by column chromatography on silica gel eluted with hexane/AcOEt (4:1) affording $U1_{Alk10}$ as a white solid (461 mg, 65%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.16 (bs, 1H, N^3H), 7.49 (s, 1H, H^6), 3.74 (t, J = 6.7 Hz, 2H, $N^1CH_2C_9H_{19}$), 3.21 (s, 1H, C≡CH), 1.8-1.6 (m, 2H, $N^1CH_2CH_2C_8H_{17}$), 1.3-1.2 (m, 14H, $N^1C_2H_4C_7H_{14}CH_3$), 0.88 (t, J = 6.7 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.6, 149.5, 148.0, 98.7, 82.1, 74.4, 49.4, 31.8, 29.7, 29.44, 29.37, 29.2, 29.1, 26.3, 22.6, 14.1.

HRMS (ESI+): Calculated for $C_{16}H_{25}N_2O_2$ $[M+H]^+$: 277.1910. Found: 277.1910.



1-hexyl-5-ethynyluracil ($U1_{Alk6}$). $U1_{Alk6}$ was prepared following *Standard Procedure B*. $U2_{Alk6}$ (0.64 g, 2.0 mmol), $Pd(PPh_3)_2Cl_2$ (27.9 mg, 0.04 mmol) and CuI (3.8 mg, 0.02 mmol) were dissolved in the THF/NEt_3 mixture (20 mL). Then TMSA (0.6 mL, 4.0 mmol) was added. The reaction was completed in 8 h. Then, TBAF·$3H_2O$ (0.6 g, 2.0 mmol) was added over a THF (15 mL) solution of the crude mixture. $U1_{Alk6}$ was purified by column chromatography on silica gel eluted with $CHCl_3$/hexane (100:1) affording $U1_{Alk6}$ as a white solid (216 mg, 49%).

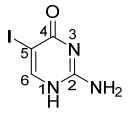
m. p. 149–151 °C

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.31 (bs, 1H, N^3H), 7.49 (s, 1H, H^6), 3.74 (t, J = 7.0 Hz, 2H, $N^1CH_2C_5H_{11}$), 3.21 (s, 1H, C≡CH), 1.8-1.6 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.4-1.2 (m, 6H, $N^1C_2H_4C_3H_6CH_3$), 0.89 (t, J = 6.7 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.9, 149.8, 148.1, 98.7, 82.1, 74.5, 49.4, 31.2, 29.0, 26.0, 22.4, 13.9.

HRMS (ESI+): Calculated for $C_{12}H_{17}N_2O_2$ $[M+H]^+$: 221.1284. Found: 221.1281.

Synthesis of the *iso*cytosine derivatives.

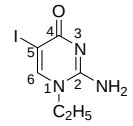


5-iodo*iso*cytosine (iC3).[188] *Iso*cytosine (1.0 g, 9.0 mmol) was suspended in a 1:1 AcOH/H_2O mixture (35 mL). The reaction mixture was stirred at 50 °C for 15 minutes. Then *N*-iodosuccinimide (2.4 g, 10.8 mmol) was added and the mixture was stirred at 100 °C for 2 h. Once completed, the reaction mixture was cooled to room temperature and the resulting cream-colored solid was filtered and washed with water, affording **iC3** as a white solid (1.7 g, 81%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.26 (bs, 1H, N^3H). 7.93 (s, 1H, H^6); 6.71 (bs, 2H, NH_2).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.5, 159.6, 156.6, 71.0.

HRMS (ESI+): Calculated for $C_4H_5N_3OI$ $[M+H]^+$: 237.9471. Found: 237.9464.



1-ethyl-5-iodo*iso*cytosine (iC2$_{Alk2}$). iC2$_{Alk2}$ was obtained following *Standard Procedure A* using **iC3** (1.7 g, 7.3 mmol), a 1.0 M solution in MeOH of Bu_4NOH (8.8 mL, 8.8 mmol) and 1-iodoethane (0.72 mL, 8.8 mmol) in DMF (30 mL). The resulting solution was stirred at room temperature for 2 h. until completion. After solvent evaporation, the crude was dissolved in $CHCl_3$ (40 mL) and washed with water (3 x 20 mL) and brine (20 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). A final recrystallization using acetonitrile yielded **iC2$_{Alk2}$** as a white solid (0.67 g, 35%).

m. p. 181–182 °C

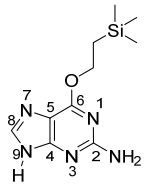
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.98 (s, 1H, H^6), 5.19 (bs, 2H, NH_2), 4.06 (q, J = 7.3 Hz, 2H, N$^1CH_2CH_3$), 1.35 (t, J = 7.3 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 159.4, 158.9, 155.2, 38.5, 29.7, 12.1.

HRMS (ESI+): Calculated for $C_6H_9N_3OI$ $[M+H]^+$: 265.9784. Found: 265.9786.

Synthesis of the guanine derivatives.

- ***Standard Procedure C*** for the carbonyl group protection of the guanine derivatives. In a double neck round-bottomed flask, equipped with a basic reflux set-up and a magnetic stirrer, activated NaH (4 eq) was placed under argon. Dry solvent (indicated in each case) was added and the mixture was stirred. Then, 2-trimethylsilylethanol (2 eq) was added dropwise and the mixture was stirred under reflux for 2.5 hours. The mixture was then allowed to cool down to room temperature and the nucleobase (1 eq) was added carefully. The resulting mixture was stirred under reflux overnight until the reaction was completed, which was monitored by TLC. Work-up and purification methods are indicated in each case.



2-amino-6-(2-(trimethylsilyl)ethoxy)purine (G7).[195] **G7** was prepared following *Standard Procedure C* using NaH (118.3 mmol, 2.8 g) and dry dioxane (125 mL). Then, 2-trimethylsilylethanol (8.5 mL, 59.2 mmol) was added dropwise. Later, 2-amino-6-chloropurine (5.0 g, 29.6 mmol) was added. The reaction was completed overnight. The reaction mixture was concentrated *in vacuo*, dissolved in water (100 mL) and washed with ether (2 x 50 mL). Acetic acid was added to the aqueous phase until pH 6 and the resulting solid was filtered. The crude solid was dried and passed through a silica plug eluted with $CHCl_3$/MeOH (10:1). A final recrystallization in MeOH/water afforded **G7** as a white solid (2.9 g, 48% yield).

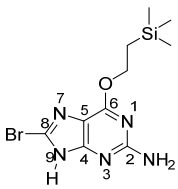
m. p. 184–185 °C

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 12.36 (bs, 1H, N^9H), 7.77 (s, 1H, H^8), 6.15 (bs, 2H, NH_2), 4.49 (t, J = 7.0 Hz, 2H, OCH_2CH$_2$TMS), 1.13 (t, J = 7.0 Hz, 2H, OCH$_2$CH_2TMS), 0.07 (s, 9H, Si(CH_3)$_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 159.9, 159.6, 156.0, 138.4, 113.0, 63.3, 17.1, -1.3.

HRMS (ESI+): Calculated for $C_{10}H_{18}N_5OSi$ $[M+H]^+$: 252.1275. Found: 252.1299.

- ***Standard Procedure D*** for the bromination reaction of the purine derivatives. To a solution of the nucleobase (1 eq) in a solvent (indicated in each case), NBS (1.1 eq) was added over a period of time (indicated in each case). The reaction mixture was stirred for a period of time (indicated in each case) at room temperature until the reaction was completed, which was monitored by TLC. Work-up and purification methods are indicated in each case.

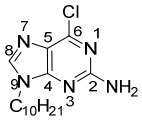


2-amino-8-bromo-6-(2-(trimethylsilyl)ethoxy)purine (G6). G6 was prepared following *Standard Procedure D*. **G7** (0.10 g, 0.4 mmol) was dissolved in acetonitrile (3 mL) and NBS (0.08 g, 0.4 mmol) was added over a period of 30 min. The reaction was completed in 3 h. The mixture was then filtered through a filter paper and the filtrate was concentrated under reduced pressure. The crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording **G6** as a yellow solid (40 mg, 35% yield).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 13.09 (bs, 1H, N^9*H*), 6.33 (bs, 2H, NH_2), 4.47 (t, *J* = 7.0 Hz, 2H, OCH_2CH$_2$TMS), 1.11 (t, *J* = 7.0 Hz, 2H, OCH$_2$CH_2TMS), 0.06 (s, 9H, Si(CH_3)$_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 159.5, 159.8, 156.1, 121.5, 114.7, 63.6, 40.4, 17.0, -1.3.

HRMS (ESI+): Calculated for $C_{10}H_{17}N_5OSiBr$ $[M+H]^+$: 330.0380. Found: 330.0395.

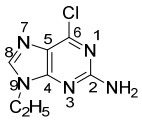


2-amino-6-chloro-9-decylpurine ($G5_{Alk10}$). $G5_{Alk10}$ was obtained following *Standard Procedure A* using 2-amino-6-chloropurine (8.0 g, 47 mmol), K_2CO_3 (7.8 g, 56.4 mmol), 1-iododecane (12 mL, 56.4 mmol) and DMF (110 mL). After solvent evaporation, the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). **$G5_{Alk10}$** was obtained as a white solid (12.4 g, 85%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.13 (s, 1H, N^8*H*), 6.87 (bs, 2H, C^{2}NH_2), 4.02 (t, *J* = 7.0 Hz, 2H, N^{9}CH_2C$_9$H$_{19}$), 1.8-1.6 (m, 2H, N^{9}CH$_2$CH_2C$_8$H$_{17}$), 1.3-1.1 (m, 14H, N^{9}C$_2$H$_4$C$_7$$H_{14}CH_3$), 0.84 (t, *J* = 7.0 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 159.7, 154.0, 149.3, 143.2, 123.4, 43.0, 31.2, 28.9, 28.8, 28.6, 28.4, 25.9, 22.0, 13.9.

HRMS (ESI+): Calculated for $C_{15}H_{25}N_5Cl$ $[M+H]^+$: 310.1793. Found: 310.1801.

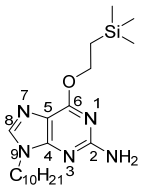


2-amino-6-chloro-9-ethyl-purine ($G5_{Alk2}$).[208] **$G5_{Alk2}$** was obtained following *Standard Procedure A* using 2-amino-6-chloropurine (3.0 g, 17.7 mmol), K_2CO_3 (2.9 g, 21.2 mmol) and 1-iodoethane (1.7 mL, 21.2 mmol) in DMF (60 mL). After solvent evaporation, the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). **$G5_{Alk2}$** was obtained as a white solid (2.6 g, 74%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.14 (s, 1H, N^8*H*), 6.91 (bs, 2H, C^{2}NH_2), 4.07 (q, *J* = 7.3 Hz, 2H, N^{9}CH_2CH$_3$), 1.36 (t, *J* = 7.3 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 159.7, 153.9, 149.3, 142.9, 123.4, 38.2, 14.9.

HRMS (ESI+): Calculated for $C_7H_9N_5Cl$ $[M+H]^+$: 198.0540. Found: 198.0542.



2-amino-9-decyl-6-(2-(trimethylsilyl)ethoxy)purine ($G4_{Alk10}$).[196] **$G4_{Alk10}$** can be obtained following *Standard Procedure A* using **G7** (1.5 g, 6.0 mmol), K_2CO_3 (1.0 g, 7.2 mmol) and 1-iododecane (1.5 mL, 7.2 mmol) in DMF (20 mL). The reaction was completed in 8 h. After solvent evaporation, the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). **$G4_{Alk10}$** was obtained as a white solid (0.97 g, 42%).

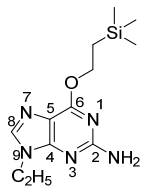
$G4_{Alk10}$ can also be obtained following *Standard Procedure C* using NaH (0.19 g, 0.7 mmol) and 2-trimethylsilylethanol (0.15 mL, 1.0 mmol) in dry THF (5 mL). Then, **$G5_{Alk10}$** (0.16 g, 0.5 mmol) was added. The reaction was completed overnight. The reaction mixture was concentrated *in vacuo*, dissolved in ethyl acetate (20 mL) and washed with water (3 x 10 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/hexane (5:1), affording **$G4_{Alk10}$** as a white solid (71 mg, 35% yield).

[208] E. Camaioni, S. Costanzi, S. Vittori, R. Volpini, K.-N. Klotz, G. Cristalli, *Bioorg. Med. Chem.* **1998**, *6*, 523–533.

^{1}H NMR (300 MHz, *DMSO*-d_6) δ(ppm) = 7.82 (s, 1H, H^8), 6.28 (bs, 2H, NH_2), 4.49 (m, 2H, OCH_2CH$_2$TMS), 3.97 (t, *J* = 7.2 Hz, 2H, N^{9}CH_2C$_9$H$_{19}$), 1.8-1.6 (m, 2H, N^{9}CH$_2$CH_2C$_8$H$_{17}$), 1.3-1.1 (m, 2H, N^{9}C$_2$H$_4$C$_7$$H_{14}CH_3$), 1.1-1.0 (m, 2H, OCH$_2CH_2$TMS), 0.84 (t, *J* = 7.1 Hz, 3H, CH_3), 0.06 (s, 9H, Si(CH_3)$_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 161.4, 159.2, 153.9, 139.1, 115.8, 64.8, 43.5, 31.8, 29.8, 29.44, 29.40, 29.2, 29.0, 26.6, 22.6, 17.6, 14.0, 1.4.

HRMS (ESI+): Calculated for $C_{20}H_{38}N_5OSi$ $[M+H]^+$: 392.2840. Found: 392.2849.



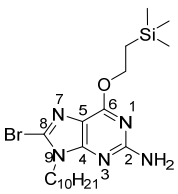
2-amino-6-(2-(trimethylsilyl)ethoxy)-9-ethylpurine (G4$_{Alk2}$).[196] **G4$_{Alk2}$** can be obtained following *Standard Procedure A* using **G7** (1.0 g, 3.9 mmol), K_2CO_3 (0.6 g, 4.7 mmol), 1-iodoethane (0.4 mL, 4.7 mmol) and DMF (20 mL). The reaction was completed in 8 h. After solvent evaporation, the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). **G4$_{Alk2}$** was obtained as a white solid (0.35 g, 32%).

G4$_{Alk2}$ can also be obtained following *Standard Procedure C* using NaH (0.86 g, 35.8 mmol) and 2-trimethylsilylethanol (2.6 mL, 17.9 mmol) in dry THF (50 mL). Then, **G5$_{Alk2}$** (1.77 g, 9.0 mmol) was added. The reaction was completed overnight. The reaction mixture was concentrated *in vacuo*, dissolved in $CHCl_3$ (50 mL), washed with ether (30 mL) and then with water (3 x 30 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/hexane (100:1), affording **G4$_{Alk2}$** as a white solid (633 mg, 25% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.60 (s, 1H, H^8), 4.77 (bs, 2H, NH_2), 4.6-4.5 (m, 2H, OCH_2CH$_2$TMS), 4.10 (q, *J* = 7.0 Hz, 2H, N^{9}CH_2CH$_3$), 1.48 (t, *J* = 7.0 Hz, 3H, CH_3), 1.3-1.2 (m, 2H, OCH$_2$CH_2TMS), 0.09 (s, 9H, Si(CH_3)$_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 162.6, 160.9, 155.0, 140.0, 116.8, 66.0, 39.7, 18.9, 16.7, 0.0.

HRMS (ESI+): Calculated for $C_{12}H_{22}N_5OSi$ $[M+H]^+$: 280.1588. Found: 280.1586.

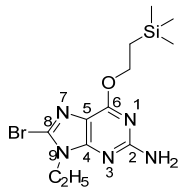


2-amino-8-bromo-9-decyl-6-(2-(trimethylsilyl)ethoxy)purine (G3$_{Alk10}$). G3$_{Alk10}$ was obtained following *Standard Procedure D* using **G4$_{Alk10}$** (0.63 g, 2.3 mmol) in CCl_4 (10 mL) and NBS (0.44 g, 2.5 mmol), added over a period of 30 min. The reaction mixture was completed 3 h. The crude mixture was filtered through a filter paper and the filtrate was concentrated under reduced pressure. The crude solid was then purified by column chromatography on silica gel eluted with $CHCl_3$/hexane (100:1), affording 0.87 g of **G3$_{Alk10}$** as a yellow solid (80% yield).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.48 (bs, 2H, NH_2), 4.48 (m, 2H, OCH_2CH$_2$TMS), 3.97 (t, *J* = 7.2 Hz, 2H, N^{9}CH_2C$_9$H$_{19}$), 1.8-1.6 (m, 2H, OCH$_2$CH_2C$_8$H$_{17}$), 1.4-1.1 (m, 14H, OC$_2$H$_4$C$_7$$H_{14}CH_3$), 1.1-1.0 (m, 2H, OCH$_2CH_2$TMS), 0.84 (t, *J* = 6.7 Hz, 3H, N^{9}C$_9$H$_{19}$CH_3), 0.06 (s, 9H, Si(CH_3)$_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 160.1, 159.0, 154.7, 135.2, 116.0, 65.0, 44.1, 31.9, 29.7, 29.5, 29.4, 29.3, 29.1, 26.5, 22.7, 17.6, 14.1, 1.4.

HRMS (ESI+): Calculated for $C_{20}H_{37}N_5OSiBr$ $[M+H]^+$: 470.1945. Found: 470.1941.



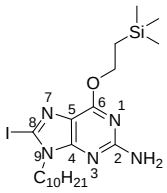
2-amino-8-bromo-6-(2-(trimethylsilyl)ethoxy)-9-ethylpurine (G3$_{Alk2}$). G3$_{Alk2}$ was obtained following *Standard Procedure D* using **G4$_{Alk2}$** (0.63 g, 2.3 mmol) in CCl_4 (10 mL) and NBS (0.44 g, 2.5 mmol), added over a period of 30 min. The reaction mixture was completed in 4 h. The crude mixture was filtered through a filter paper and the filtrate was concentrated under reduced pressure. The crude solid was then purified by column chromatography on silica gel eluted with $CHCl_3$/hexane (100:1), affording 0.60 g of **G3$_{Alk2}$** (72% yield).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.48 (bs, 2H, NH_2), 4.48 (m, 2H, OCH_2CH$_2$TMS), 4.02 (q, *J* = 7.1 Hz, 2H, N^{9}CH_2CH$_3$), 1.26 (t, *J* = 7.2 Hz, 3H, CH_3), 1.11 (m, 2H, OCH$_2$CH_2TMS), 0.06 (s, 9H, Si(CH_3)$_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 159.8, 159.2, 154.5, 123.7, 114.1, 63.8, 55.6, 17.0, 14.4, -1.3.

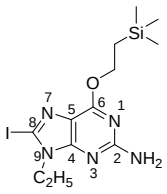
HRMS (ESI+): Calculated for $C_{12}H_{21}N_5OSiBr$ $[M+H]^+$: 358.0693. Found: 358.0693.

- ***Standard Procedure E*** for the iodination reaction of the guanine derivatives. To a solution of the nucleobase (1 eq) in THF (volume indicated in each case) was added a solution of LDA freshly prepared from diisopropylamine (7.5 eq) and *n*BuLi (2.5 M, 5.0 eq) in THF (volume indicated in each case) at -78 °C. After 4 h, a solution of I_2 (3.0 eq) in THF (volume indicated in each case) was added and the mixture was stirred at -78 °C until the reaction was completed, which was monitored by TLC. The mixture was left to warm up to 0 °C, followed by hydrolysis with NH_4Cl (sat). The water phase was separated and extracted with CH_2Cl_2. The combined organic layers were concentrated under reduced pressure, dissolved in CH_2Cl_2 and washed with $Na_2S_2O_3$ (sat.), $NaHCO_3$ (sat.) and brine. The organic layer was then dried over $MgSO_4$ and concentrated *in vacuo*. Purification methods are indicated in each case.



2-amino-8-iodo-9-decyl-6-(2-(trimethylsilyl)ethoxy)purine ($G2_{Alk10}$). $G2_{Alk10}$ was obtained following *Standard Procedure E*, LDA was prepared from diisopropylamine (0.28 mL, 2.0 mmol) and a 2.5 M *n*BuLi solution in THF (0.52 mL, 1.3 mmol) in THF (3 mL). The mixture was stirred at 0 °C for 45 min, cooled at -78 °C and then a solution of **$G4_{Alk10}$** (0.10 g, 0.3 mmol) in THF (5 mL) was added *via cannula*. The resulting mixture was stirred at -78 °C for 5 h. A solution of I_2 (15 mg, 0.8 mmol) in THF (10 mL) was added. The reaction was completed in 2 h. Then, NH_4Cl (sat.) was added to the mixture and the system was allowed to slowly reach room temperature. The phases were separated and the aqueous phase was extracted with CH_2Cl_2 (3 x 4 mL). The combined organic layers were concentrated under reduced pressure, dissolved in CH_2Cl_2, washed with $Na_2S_2O_3$ (sat.) (1 x 5 mL), $NaHCO_3$ (sat.) (5 mL) and brine (5 mL). The organic layer was dried over $MgSO_4$ and the solvent was evaporated. The resulting solid was purified by column chromatography eluted with $CHCl_3$/cyclohexane (100:1), affording **$G2_{Alk10}$** as a yellow solid (74 mg, 55%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 4.78 (bs, 2H, NH_2), 4.57 (m, *J* = 8.1 Hz, 2H, OCH_2CH$_2$TMS), 4.03 (t, *J* = 7.1 Hz, 2H, N^{9}CH_2C$_9$H$_{19}$), 1.83 (m, 2H, OCH$_2$CH_2C$_8$H$_{17}$), 1.38-1.17 (m, 16H, OC$_2$H$_4$C$_7$$H_{14}CH_3$, OCH$_2CH_2$TMS), 0.87 (t, *J* = 7.1 Hz, 3H, N^{9}C$_9$H$_{19}$CH_3, 0.09 (s, 9H, Si(CH_3)$_3$).

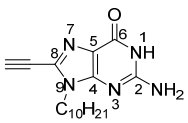


2-amino-8-iodo-9-decyl-6-(2-(trimethylsilyl)ethoxy)purine ($G2_{Alk2}$). $G2_{Alk2}$ was obtained following *Standard Procedure E*. LDA was prepared from diisopropylamine (8.8 mL, 62.2 mmol) and a 2.5 M *n*BuLi solution in THF (16.6 mL, 41.5 mmol) in THF (90 mL). The mixture was stirred at 0 °C for 45 min, cooled at -78 °C and then a solution of **$G4_{Alk2}$** (2.33 g, 8.3 mmol) in THF (100 mL) was added *via cannula*. The resulting mixture was stirred at -78 °C for 5 h. A solution of I_2 (6.3 g, 24.9 mmol) in THF (100 mL) was added. The reaction was completed in 3 h. Then, NH_4Cl (sat.) was added to the mixture and the system was allowed to slowly reach rt. The phases were separated and the aqueous phase was extracted with CH_2Cl_2 (3 x 20 mL). The combined organic layers were concentrated under reduced pressure, dissolved in CH_2Cl_2, washed with $Na_2S_2O_3$ (sat.) (1 x 50 mL), $NaHCO_3$ (sat.) (50 mL) and brine (50 mL). The organic layer was dried over $MgSO_4$ and the solvent was evaporated. The resulting solid was purified by column chromatography eluted with $CHCl_3$/MeOH (100:1), affording **$G2_{Alk2}$** as a yellow solid (2.4 g, 72%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 4.84 (bs, 2H, NH_2), 4.53 (m, 2H, OCH_2CH$_2$TMS), 4.07 (t, *J* = 7.1 Hz, 2H, N^{9}CH_2CH$_3$), 1.36 (t, *J* = 7.1 Hz, 3H, N^{9}CH$_2$CH_3), 1.21 (m, 2H, OCH$_2$CH_2TMS), 0.07 (s, 9H, Si(CH_3)$_3$) ppm.

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.0, 160.4, 155.9, 118.0, 100.3, 65.1, 41.2, 18.3, 16.1, 0.0.

HRMS (ESI+): Calculated for $C_{12}H_{21}N_5OSiI$ $[M+H]^+$: 406.0554. Found: 406.0551.



9-decyl-8-ethynylguanine ($G1_{Alk10}$). $G1_{Alk10}$ was prepared following *Standard Procedure B*. **$G3_{Alk10}$** (2.3 g, 4.9 mmol), $Pd(PPh_3)_2Cl_2$ (68.6 mg, 0.10 mmol) and CuI (9.3 mg, 0.05 mmol) were dissolved in the THF/NEt_3 mixture (20 mL). Then TMSA (1.80 mL, 9.8 mmol) was added. The reaction was completed in 12 h. Then, the mixture was concentrated under reduced pressure, dissolved in THF (30 mL) and TBAF·3H_2O (3.4 g, 10.8 mmol) was

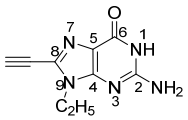
added. After solvent evaporation, the dark brown oil was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (30:1). **G1$_{Alk10}$** was obtained as a light yellow solid (1.1 g, 70%).

m. p. > 250 °C

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.68 (bs, 1H, N^1*H*), 6.59 (bs, 2H, NH_2), 4.70 (s, 1H, C≡C*H*), 3.97 (t, *J* = 6.9 Hz, 2H, N^{9}CH_2C$_9$H$_{19}$), 1.8-1.6 (m, 2H, N^{9}CH$_2$CH_2C$_8$H$_{17}$), 1.3-1.1 (m, 14H, N^{9}C$_2$H$_4$C$_7$*H*$_{14}$CH$_3$), 0.83 (t, *J* = 6.9 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 156.2, 84.8, 73.6, 42.7, 31.2, 28.85, 28.78, 28.6, 28.4, 25.8, 22.0, 13.9.

HRMS (ESI+): Calculated for $C_{17}H_{26}N_5O$ $[M+H]^+$: 316.2131. Found: 316.2146.



9-ethyl-8-ethynylguanine (G1$_{Alk2}$). G1$_{Alk2}$ was prepared following *Standard Procedure B*. **G3$_{Alk2}$** (0.57 g, 1.6 mmol), $Pd(PPh_3)_2Cl_2$ (22.0 mg, 0.03 mmol) and CuI (3.0 mg, 0.02 mmol) were dissolved in THF/NEt_3 (15 mL). Then TMSA (0.6 mL, 3.2 mmol) was added. The reaction was completed in 8 h. Then, the mixture was concentrated under reduced pressure, dissolved in THF (20 mL) and TBAF·3H_2O (1.1 g, 3.5 mmol) was added. After solvent evaporation, the dark brown oil was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (10:1). **G1$_{Alk2}$** was obtained as a light yellow solid (270 mg, 83%).

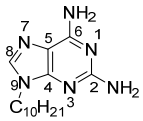
m. p. > 250 °C

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.71 (bs, 2H, NH^1), 6.61 (bs, 2H, NH_2), 4.71 (s, 1H, C≡C*H*), 4.01 (q, *J* = 7.0 Hz, 2H, N^{9}CH_2CH$_3$), 1.27 (t, *J* = 7.0 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 156.2, 154.2, 150.7, 128.8, 116.5, 84.8, 73.4, 37.7, 14.8.

HRMS (ESI+): Calculated for $C_9H_{10}N_5O$ $[M+H]^+$: 204.0879.Found: 204.0878.

Synthesis of the 2-aminoadenine derivatives.

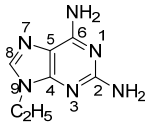


2,6-diamino-9-decylpurine (A4$_{Alk10}$). A4$_{Alk10}$ was obtained following *Standard Procedure A* using 2,6-diaminopurine (5.0 g, 33 mmol), K_2CO_3 (5.5 g, 40 mmol) and 1-iododecane (8.6 mL, 40 mmol) in DMF (100 mL). The reaction was completed overnight. After solvent evaporation, the crude was dissolved in $CHCl_3$ (200 mL) and washed with water (3 x 100 mL) and brine (100 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). **A4$_{Alk10}$** was obtained as a white solid (6.7 g, 69%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 7.69 (s, 1H, N^8*H*), 6.61 (bs, 2H, C^{6}NH_2), 5.75 (bs, 2H, C^{2}NH_2), 3.92 (t, *J* = 6.7 Hz, 2H, N^{9}CH_2C$_9$H$_{19}$), 1.8-1.6 (m, 2H, N^{9}CH$_2$CH_2C$_8$H$_{17}$), 1.3-1.1 (m, 14H, N^{9}C$_2$H$_4$C$_7$*H*$_{14}$CH$_3$), 0.84 (t, *J* = 6.5 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 160.2, 156.1, 151.2, 137.6, 113.2, 42.4, 31.3, 29.3, 29.0, 28.8, 28.7, 28.6, 26.1, 22.1, 14.0.

HRMS (ESI+): Calculated for $C_{15}H_{27}N_6$ $[M+H]^+$: 291.2291. Found: 291.2302.

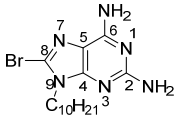


2,6-diamino-9-ethylpurine (A4$_{Alk2}$). A4$_{Alk2}$ was obtained following *Standard Procedure A* using 2,6-diaminopurine (3.0 g, 20 mmol), K_2CO_3 (3.3 g, 24 mmol) and 1-iodoethane (2 mL, 24 mmol) in DMF (150 mL). The reaction was completed overnight. The crude mixture was concentrated under reduced pressure and the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (10:1), affording 2.4 g of **A4$_{Alk2}$** (67% yield) as a white solid.

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 7.71 (s, 1H, N^8*H*), 6.64 (bs, 2H, C^{6}NH_2), 5.77 (bs, 2H, C^{2}NH_2), 3.97 (q, *J* = 7.3 Hz, 2H, N^{9}CH_2CH$_3$), 1.32 (t, *J* = 7.3 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 160.2, 156.0, 151.3, 137.0, 113.3, 37.3, 15.3.

HRMS (ESI+): Calculated for $C_7H_{11}N_6$ $[M+H]^+$: 179.1039. Found: 179.1038.



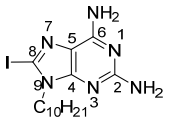
2,6-diamino-8-bromo-9-decylpurine (A3$_{Alk10}$). A3$_{Alk10}$ was obtained following *Standard Procedure D* using **A4$_{Alk10}$** (4.9 g, 16.9 mmol) in acetonitrile (80 mL) and NBS (3.6 g, 18.6 mmol), which was added over a period of 4 h. The reaction was completed overnight. Then, the crude mixture was concentrated under reduced pressure and the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (30:1), affording 0.8 g of **A3$_{Alk10}$** as an ochre solid (13% yield).

m. p. 140–142 ^oC

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.77 (bs, 2H, C^6NH_2), 5.87 (bs, 2H, C^2NH_2), 3.90 (t, J = 7.1 Hz, 2H, $N^9CH_2C_9H_{19}$), 1.68 (m, 2H, $N^9CH_2CH_2C_8H_{17}$), 1.3-1.1 (m, 14H, $N^9C_2H_4C_7H_{14}CH_3$), 0.82 (t, J = 7.1 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 160.3, 155.0, 152.7, 121.0, 113.5, 43.0, 31.2, 28.8, 28.7, 28.6, 28.5, 25.9, 22.0, 13.9.

HRMS (ESI+): Calculated for $C_{15}H_{26}BrN_6$ $[M+H]^+$: 369.1396. Found: 369.1411.

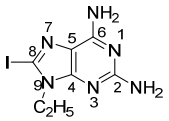


2,6-diamino-9-decyl-8-iodopurine (A2$_{Alk10}$).[174] Iodine (2.3 g, 9.3 mmol) was added to a suspension of H_5IO_6 (1.4 g, 6.2 mmol) in MeOH (10 mL) at room temperature. After 10 min, a solution of **A2$_{Alk10}$** (1.8 g, 6.2 mmol) in MeOH (10 mL) was added at rt. The reaction mixture was stirred at 70 ^oC for 5 h. After completion, the mixture was let to cool down to rt and $Na_2S_2O_5$ (sat) (10 mL) was added. The mixture was extracted in CH_2Cl_2 (30 mL) and washed with water (20 mL) and brine (20 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). **A2$_{Alk10}$** was obtained as a white solid (1.3 g, 50%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.74 (bs, 2H, C^6NH_2), 5.81 (bs, 2H, C^2NH_2), 3.86 (t, J = 7.3 Hz, 2H, $N^9CH_2C_9H_{19}$), 1.7-1.6 (m, 2H, $N^9CH_2CH_2C_8H_{17}$), 1.3-1.1 (m, 14H, $N^9C_2H_4C_7H_{14}CH_3$), 0.84 (t, J = 6.9 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 160.2, 154.8, 152.8, 116.2, 95.8, 44.3, 31.3, 29.1, 28.88, 28.87, 28.7, 28.6, 26.0, 22.1, 13.9.

HRMS (ESI+): Calculated for $C_{15}H_{26}IN_6$ $[M+H]^+$: 417.1258. Found: 417.1275.



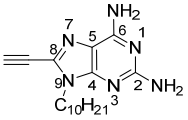
2,6-diamino-9-ethyl-8-iodopurine (A2$_{Alk2}$). A2$_{Alk2}$ (1.0 g, 5.6 mmol), iodine (1.3 g, 2.1 mmol) and H_5IO_6 (1.3 g, 5.6 mmol) were stirred in DMF (30 mL) at 70 ^oC for 48 h. After completion, the reaction mixture was concentrated *in vacuo*. The crude was dissolved in $CHCl_3$ (50 mL) and washed with $Na_2S_2O_5$ (sat) (40 mL), water (40 mL) and brine (40 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo*. The crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (10:1). **A2$_{Alk2}$** was obtained as a white solid (0.6 g, 37%).

m. p. 242–243 ^oC

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.77 (bs, 2H, C^6NH_2), 5.87 (bs, 2H, C^2NH_2), 3.92 (q, J = 7.4 Hz, 2H, $N^9CH_2CH_3$), 1.23 (t, J = 7.0 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 160.3, 154.9, 152.5, 116.3, 95.4, 38.8, 15.0.

HRMS (ESI+): Calculated for $C_7H_{10}N_6I$ $[M+H]^+$: 305.0006. Found: 305.0001.



2,6-diamino-9-decyl-8-ethynylpurine (A1$_{Alk10}$). A1$_{Alk10}$ was prepared following *Standard Procedure B*. **A2$_{Alk10}$** (0.34 g, 0.8 mmol), $Pd(PPh_3)_2Cl_2$ (11.4 mg, 0.02 mmol) and CuI (1.5 mg, 0.01 mmol) were dissolved in the THF/NEt_3 mixture (13 mL). Then TMSA (0.3 mL, 1.6 mmol) was added. The reaction was completed in 12 h. The mixture was concentrated under reduced pressure, dissolved in THF (15 mL) and TBAF·$3H_2O$ (0.3 g, 0.9 mmol) was added. **A1$_{Alk10}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH: (20:1) affording **A1$_{Alk10}$** as a white solid (130 mg, 51%).

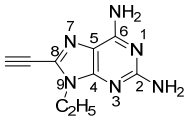
A1$_{Alk10}$ can also been obtained following *Standard Procedure B* with **A3$_{Alk10}$** (0.91 g, 2.5 mmol), $Pd(PPh_3)_2Cl_2$ (36.1 mg, 0.05 mmol) and CuI (4.9 mg, 0.02 mmol) were dissolved in the THF/NEt_3 mixture (10 mL). Then TMSA (0.9 mL, 5.0 mmol) was added. The reaction was completed in 12 h. The mixture was concentrated under reduced pressure, dissolved in THF (15 mL) and TBAF·$3H_2O$ (0.9 g, 2.8 mmol) was added. **A1$_{Alk10}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1) affording **A1$_{Alk10}$** as a white solid (130 mg, 82%).

m. p. 144–145 °C

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 6.86 (bs, 2H, C^6NH_2), 5.94 (bs, 2H, C^2NH_2), 4.69 (s, 1H, C≡C*H*), 3.98 (t, *J* = 6.9 Hz, 2H, $N^9CH_2C_9H_{19}$), 1.71 (m, 2H, $N^9CH_2CH_2C_8H_{17}$), 1.4-1.1 (m, 14H, $N^9C_2H_4C_7H_{14}CH_3$), 0.84 (t, *J* = 6.9 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.2, 156.1, 151.4, 129.0, 113.1, 84.7, 74.0, 42.4, 31.3, 28.95, 28.90, 28.7, 28.6, 26.0, 22.1, 13.9.

HRMS (ESI+): Calculated for $C_{17}H_{27}N_6$ $[M+H]^+$: 315.2291. Found: 315.2289.



2,6-diamino-9-ethyl-8-ethynylpurine (A1$_{Alk2}$). **A1$_{Alk2}$** was prepared following *Standard Procedure B*. **A2$_{Alk2}$** (0.45 g, 1.5 mmol), $Pd(PPh_3)_2Cl_2$ (20.6 mg, 0.03 mmol) and CuI (2.8 mg, 0.01 mmol) were dissolved in the THF/NEt_3 mixture (35 mL). Then, TMSA (0.4 mL, 3.0 mmol) was added. The reaction was completed in 12 h. The mixture was concentrated under reduced pressure, dissolved in THF (35 mL) and TBAF·$3H_2O$ (0.55 g, 1.65 mmol) was added. **A1$_{Alk2}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH: (10:1) affording **A1$_{Alk2}$** as a white solid (146 mg, 48%).

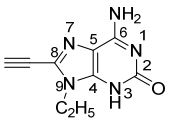
m. p. > 250 °C

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 5.41 (bs, 2H, C^6NH_2), 4.77 (bs, 2H, C^2NH_2), 4.18 (q, *J* = 7.2 Hz, 2H, $N^9CH_2CH_3$), 3.42 (s, 1H, C≡C*H*), 1.42 (t, *J* = 7.2 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.2, 156.1, 151.1, 128.4, 113.1, 84.7, 73.9, 37.3, 14.9.

HRMS (ESI+): Calculated for $C_9H_{11}N_6$ $[M+H]^+$: 203.1039. Found: 203.1040.

Synthesis of the *iso*guanine derivative.



9-ethyl-8-ethynyl*iso*guanine (iG1$_{Alk2}$).[199a] **iG1$_{Alk2}$** was prepared from **A2$_{Alk2}$** (0.15 g, 0.5 mmol) following the same procedure than **A1$_{Alk2}$** but, before TMS-deprotection, the crude mixture after the Sonogashira coupling was suspended in a 1:1 THF/H_2O mixture (18 mL). Then, a solution of $NaNO_2$ (0.15 g, 2.2 mmol) in H_2O (2 mL) was added. The mixture was heated at 50 °C and a 1:1 AcOH/H_2O mixture (0.5 mL) was slowly added. The reaction mixture was stirred at 50 °C for 2 h. Once completed, the mixture was let to cool down to rt and NH_3 was added until pH 8. The reaction mixture was concentrated under reduced pressure. Then, K_2CO_3 (0.15 g, 1.1 mmol) was added over a MeOH (10 mL) solution of the resulting crude solid. The reaction mixture was stirred at rt for 1 h. until completion. After solvent evaporation the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (20:1). **iG1$_{Alk2}$** was obtained as a pale yellow solid (81 mg, 73%).

m. p. > 250 °C

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.0 (bs, 1H, N^3H), 7.90 (bs, 2H, NH_2), 4.80 (s, 1H, C≡C*H*), 3.97 (q, *J* = 7.2 Hz, 2H, $N^9CH_2CH_3$), 1.26 (t, *J* = 7.2 Hz, 3H, CH_3).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 130.3, 85.5, 79.2, 73.2, 56.1, 45.9, 37.5, 14.7, 8.7.

HRMS (ESI+): Calculated for $C_9H_{10}N_5O$ $[M+H]^+$: 204.0879. Found: 204.0879.

Chapter 2.

Molecular Self-Assembly at the Solid-Liquid Interface.

The STM experiments at the solid-liquid interface described in this Chapter were carried out at the Division of Molecular Imaging and Photonics of the Department of Chemistry of the *Katholieke Universiteit Leuven* (Belgium) under the supervision of Prof. Steven De Feyter during two predoctoral short stays of three and two months, respectively.

2.1. Introduction. Molecular imaging by STM at the solid-liquid interface.

It has been seen in the second part of the Introduction of this Thesis how important it is to drive the ordering of conjugated oligomers and polymers on flat solid surfaces with the aim of producing highly ordered and thermodynamically stable supramolecular architectures for nanotechnological applications. Self-assembly appears as a effective way to form ordered monolayers at interfaces, both under UHV at low temperatures and under ambient conditions, as well as at the solid-liquid and solid-air interfaces. STM arises as the most interesting and powerful tool to study and understand the arrangement of the molecules on solid substrates in such conditions.

The invention of the scanning tunneling microscope in the early 1980s[209] opened new ways to investigate surface phenomena on a (sub)nanometer scale thanks to the very localized nature of the probing. In STM, a sharp metallic tip, usually made from a platinum-iridium alloy or tungsten metal, is brought very close (3–5 Å) to a conductive substrate and an overlap occurs between tip and sample wave functions, decaying exponentially into the junction gap. By applying a voltage between both electrodes, this vacuum tunneling establishes a small tunnel current (I_t), within the nano-ampere range, through a classically impenetrable barrier. The direction of the tunneling depends on the bias polarity (V_{bias}). Since the tunnel current depends exponentially on the distance (z) between the tip and the surface, the individual atoms in the surface will give rise to current variations as the tip is scanned across the corrugated surface. That is, the tunneling current tends to increase (decrease) as the separation between the tip and the sample decreases (increases). When an excellent control of the distance z between the probe and the surface is obtained, a very high resolution on atomically flat conductive substrates can be achieved. The contrast in STM images reflects both topography and electronic effects. There are mainly two modes of operation: the constant-height mode and the constant-current mode. In the former, the absolute vertical position of the probe remains constant during the line-by-line scanning. In the latter, the signal of the probe is kept constant through readjusting the vertical position of the probe or the sample. The vertical position of the probe is plotted as a function of the lateral position.[210] Generally, the constant-current mode yields more reliable information about topographic height, whereas the constant-height mode allows

[209] a) G. Binnig, H. Rohrer, C. Gerber, E. Weibel, *Phys. Rev. Lett.* **1982**, *49*, 57–61; b) G. Binnig, H. Rohrer, *Helv. Phys. Acta* **1982**, *55*, 726–735.
[210] S. De Feyter, F. De Schryver, *Top. Curr. Chem.* **2005**, *258*, 205–255.

for higher scanning frequencies. Normally, the STM is operated in the constant-current mode in which the actual I_t is compared with a preset constant value (I_0), typically 0.5–5 nA, in a feedback circuit. The feedback signal, proportional to the difference between I_t and I_0, provides a correction voltage to the *z* transducer and thus causes the distance *z* between the tip and the surface to change when an atom or, in more general terms, a protrusion is traversed. Recording the feedback signal or *z* voltage as a function of the lateral tip position during raster scanning, yields a map of the surface topography. Finally, in order to image molecules adsorbed on the atomically flat conductive substrate, the molecules have to be immobilized on the imaging time scale.

Adsorption is the adhesion of molecules from a gas or a liquid to a surface. This process creates a film of the *adsorbate* on the surface of the *adsorbent*. It is a surface-based process where depending on the bonding nature it can be considered chemisorption (covalent bonding) (Figure 52) or physisorption (van der Waals and electrostatic interactions).

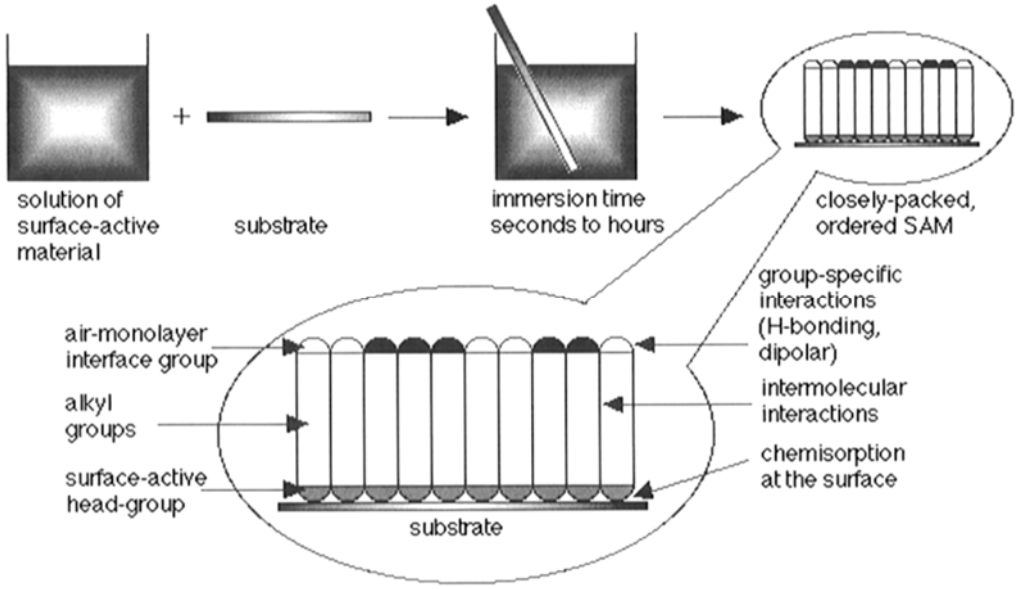


Figure 52. SAMs are formed by immersing a substrate into a solution of the surface-active material. The driving force for the spontaneous formation of the 2D assembly includes chemical bond formation of molecules with the surface (chemisorption) and lateral intermolecular interactions.[211]

Chemisorption is mainly represented by the self-assembly of surfactant molecules at surfaces, where both head and tail groups are tailored for the formation of monolayers and study of different phenomena like ordering, growth, wetting, adhesion, lubrication and corrosion. Monolayers of alkanethiolates on crystalline gold surfaces are probably the most

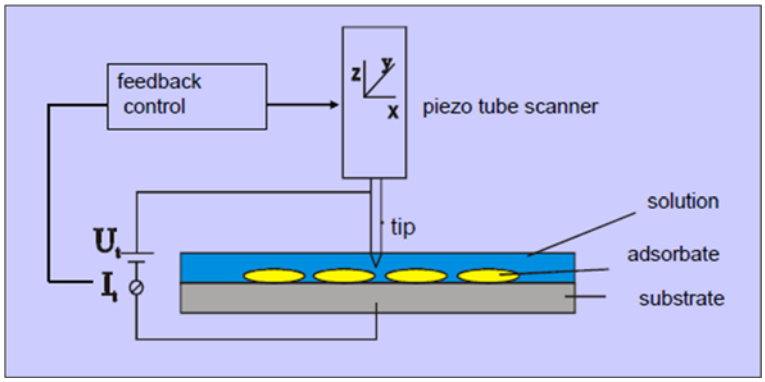
[211] A. Ulman, *Chem. Rev.* **1996**, *96*, 1533–1554.

studied chemisorbed self-assembled monolayers (SAMs) to date, as organosulfur compounds coordinate very strongly to transition metal surfaces (Figure 52).[212]

In the case of physisorption, the adsorbate is able to diffuse on the surface until an equilibrium is established and an adlayer is formed. This process is reversible and the molecules can be desorbed. The adsorption process is energetically more favorable when the molecules interact with neighboring molecules on the surface. The physisorption of molecules on a surface is enthalpically driven and the adsorbate density will be maximized. In this Thesis the focus has been put on the formation of one atom thick single monolayers physisorbed on the surface.

Monolayer formation and molecular imaging by STM can be performed under different conditions. Under ambient conditions, high-quality STM images can also be accomplished. SAMs form spontaneously by physisorption at the solid-liquid interface (Scheme 14) and if the molecules are laterally immobilized by adsorbate-adsorbate and adsorbate-substrate interactions, stable imaging is achieved. In contrast to the solid-air interface, at the solid-liquid interface the adsorbed molecules are in equilibrium with those dissolved in the solution, leading to adsorption-desorption dynamics. In the adsorption of the molecules on the surface, concentration is one of the most important factors. At such liquid-solid interfaces, the process occurs as adsorbed molecules replace liquid molecules. The system will try to reach the thermodynamic equilibrium by minimizing its total free energy. This means that the adsorbate density will be maximized (minimum enthalpy value) and the loss of degrees of freedom will be minimized (maximum entropy value). The adsorption of larger and more rigid molecules will be consequently favored. However, the thermodynamically most stable structures are not always formed and imaged on the surface as kinetics play an important role as well. This can sometimes be avoided by annealing the samples. The self-assembly process has, in these conditions, an intermediate behavior as in solution and in the crystal.

[212] a) L. H. Dubois, R. G. Nuzzo, *Ann. Phys. Chem.* **1992**, *43*, 437–463; b) C. D. Bain, G. M. Whitesides, *Adv. Mater.* **1989**, *1*, 506–516; c) J. P. Folkers, J. A. Zerkowski, P. E. Laibinis, C. T. Seto, G. M. Whitesides, *Supramolecular architecture: Synthetic Control in Thin Films and Solids* (Ed.: T. Bein), *ACS Symposium Series 499;* American Chemical Society, Washington DC, **1992**, pp. 10–23; d) T. R. Lee, P. E. Laibinis, J. P. Folkers, G. M. Whitesides, *Pure Appl. Chem.* **1991**, *63*, 821–828; e) G. M. Whitesides, G. S. Ferguson, *Chemtracts-Org. Chem.* **1988**, *1*, 171–187.



Scheme 14. Schematic view of a basic STM setup for measurements at the solid-liquid interface.

Our contribution:

In this Thesis we will evaluate different molecular STM imaging conditions (see Background and Objectives section) and, in this particular Chapter, we will focus on STM at the solid-liquid interface (Scheme 14). We will make use of this potent tool to address the transfer, from solution to the surface, of the supramolecular chemistry of molecules endowed with a preorganized structure for ring closure. If self-assembled macrocycles can be preserved on the surface, the versatility of our strategy may allow us to precisely tune the geometry, chirality and function of ordered nanoporous networks. In a second step, we aim to study the selective recognition of molecular guests as a function of their structural match with the tailored pores.

2.2. Results and Discussion.

The monomers discussed in this section have been prepared *via* a general Pd-catalyzed Sonogashira coupling route from the molecular fragments prepared in the previous Chapter 1 (Figure 53). In this general sequence, "base 1" is most frequently a pyrimidine derivative and the purine is incorporated later as "base 2", since its manipulation is more delicate for solubility reasons and it is typically more valuable. When a mono-coupling product is targeted, more equivalents of the central block than equivalents of the ethynylated nucleobase derivative are used in the first step of the synthetic route, in order to maximize the yield of the reaction. However, small amounts of the di-substituted product, as well as of the homo-coupling product, are always observed (see also Scheme 3). On the contrary, an excess of the base is employed when the symmetrically doubly-coupled monomer is desired. The synthetic details for the preparation of each monomer can be found at the Experimental Section (**2.4.**) of this Chapter.

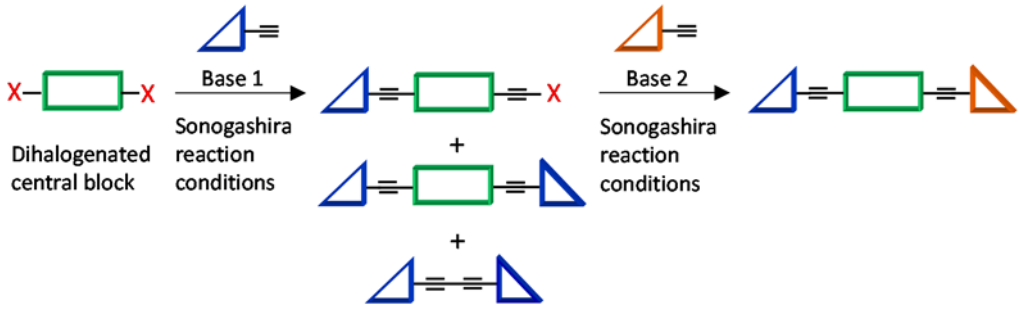


Figure 53. General synthetic pathway towards final target monomers.

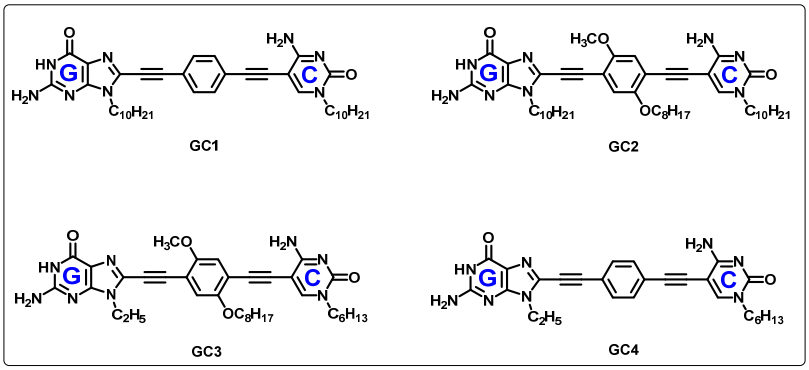
A series of molecules we anticipated convenient for the investigation of different aspects (see Objectives section) have been prepared in this Thesis. Some of them remain, nonetheless, not studied at the moment. The purpose of each one of these target molecules that have not been examined yet will be gradually explained in this section.

2.2.1. One-component Systems: toward cyclic tetramer porous networks.

Molecular Design.

In order to progress in the study of the formation and the behavior of self-assembled monolayers based on our building blocks, the simplest systems were considered first: ditopic monomers carrying complementary bases at the edges so that they form cyclic tetramers composed of 1 component. The strongest and more preorganized G–C Watson-Crick base-pair has been chosen, so that we obtain a higher chelate predisposition toward cyclic arrangements (see Background of the Group). In that way, the study of the assembly would be easier than with other more complex or weaker systems. Therefore molecules

GC1-**GC4** were prepared (Scheme 15). Furthermore, the work presented here was performed using HOPG, which is often used as substrate at the solid-liquid interface because it is electrically conductive, atomically flat, easy to clean, inert and stable in the air. Metal substrates, especially Au (111), tend to interact strongly with aromatic molecules[213] and due to the higher diffusion barriers, it is more complicated to control organization and achieve long-range order through physisorption of molecules, compared to HOPG. In this way, HOPG appeared to be the most convenient substrate for this investigation, in terms of granting a strong adsorption but maintaining an adequate mobility of the molecules.



Scheme 15. Chemical structure of the considered GC monomer family.

All molecules are composed by an arene central block *p*-disubstituted with the G and C moieties that carry alkyl chains. The choice of these alkyl chains was made based on three points: (1) to provide our molecular components with sufficient solubility to ease their manipulation; (2) to stabilize the 2D network formed through van der Waals secondary interactions *via* interdigitation of the alkyl chains; and (3) to endow the monomer with good affinity for the HOPG substrate. Long alkyl chains show strong affinity for HOPG as they usually align parallel to each other on the substrate because of the similarity between the values of the HOPG honeycomb period and the methyl-methyl distance in an alkyl chain, which favors strong [CH…π] interactions.[164] However, they should not be very long either. It is well-known that numerous molecules present different polymorphism on the surface. Different studies carried out for understanding and controlling the phase behavior showed that by increasing or decreasing the chain length, the equilibrium between densely-packed networks and porous 2D-assemblies can be shifted. The longer the alkyl chains, the more the packing is dominated by them and the steric hindrance of the aromatic cores becomes less important. By increasing the length of the alkyl tails, the porous structure probability decreases.[214] As porous structures wanted to be favored and H-bonding was required to be

[213] a) A. Bhattarai, U. Mazur, K. W. Hipps, *J. Am. Chem. Soc.* **2014**, *136*, 2142–2148; b) T. Balandina, K. Tahara, N. Sändig, M. O. Blunt, J. Adisoejoso, S. Lei, F. Zerbetto, Y. Tobe, S. De Feyter, *ACS Nano* **2012**, *6*, 8381–8389.
[214] a) P. Wu, Q. Zeng, S. Xu, C. Wang, S. Yin, C. L. Bai, *ChemPhysChem* **2001**, *2*, 750–754; b) F. Charra, J. Cousty, *Phys. Rev. Lett.* **1998**, *80*, 1682–1685; c) L. Askadskaya, C. Boeffel, J.P. Rabe, *Bunsenges. Phys. Chem.* **1993**, *97*, 517–521; d) S. Ito,

the primary interaction in directing the self-assembly, the decision was made on equipping our monomers with alkyl chains shorter than $–C_{12}$, so that van der Waals interactions would not completely rule over H-bonding.[215]

Monomer **GC1** carries a simple unsubstituted benzene ring as central block and the self-assembling units, G and C, are both functionalized with long decyl chains. The second molecular candidate shares with **GC1** a common design, except that **GC2** presents a substituted phenyl ring with a methoxy group on one side and an octyloxy group on the other. **GC3** carries the same unsymmetric substituted central block than **GC2** but each nucleobase is endowed with shorter alkyl chains. **GC4** has an intermediate structure between **GC1** and **GC3**. This monomer is formed by a phenyl ring *p*-disubstituted with G and C nucleobases functionalized with short $–C_2$ and $–C_6$ alkyl chains, respectively. These structural changes were progressively incorporated in the new designs taking feedback from the experimental data into account.

With these molecules, the self-assembly process on the surface wanted to be addressed. The primary goal was to favor the formation of cyclic tetramers over open oligomers (or other structures), when the molecules are concentrated on surfaces. As it was introduced before, the intramolecular event that shifts the equilibria toward cyclization is only relevant at low concentrations. When the molecules are concentrated onto a substrate, intra- and intermolecular binding events are compensated and closed H-bonded species do not necessarily dominate over open oligomers. Therefore, this primary task is the most challenging one, as it is not evident how to prioritize cyclic species over polymeric arrays, even if the molecule is preorganized for ring closure. The next objective will be to identify the secondary interactions that stabilize cyclic tetramer 2D networks. Prior to our studies we hypothesized that several networks could be in principle formed with **GC1**. In Figure 54 we show some proposed association modes for the GC monomers in two dimensions. If the main structure is a cyclic tetramer, we could devise the formation of tetramer networks stabilized by van der Waals interactions between the alkyl tails. However, another secondary interaction may lead to denser tetramer networks: self-complementary H-bonding interactions between the groups that do not participate in Watson-Crick bonding. Notice that both of these secondary interactions may also lead to networks of open oligomers/polymers. Finally, any void space left between cyclic tetramers that could act as an unspecific host cavity for guest molecules should be filled.

M. Wehmeier, J. D. Brand, C. Kübel, R. Epsch, J. P. Rabe, K. Müllen, *Chem Eur. J.* **2000**, *6*, 4327–4342; e) E. Mena-Osteriz, *Adv. Mater.* **2002**, *14*, 609–616.

[215] P. N. Dickerson, A. M. Hibberd, N. Oncel, S. L. Bernasek, *Langmuir* **2010**, *26*, 18155–18161.

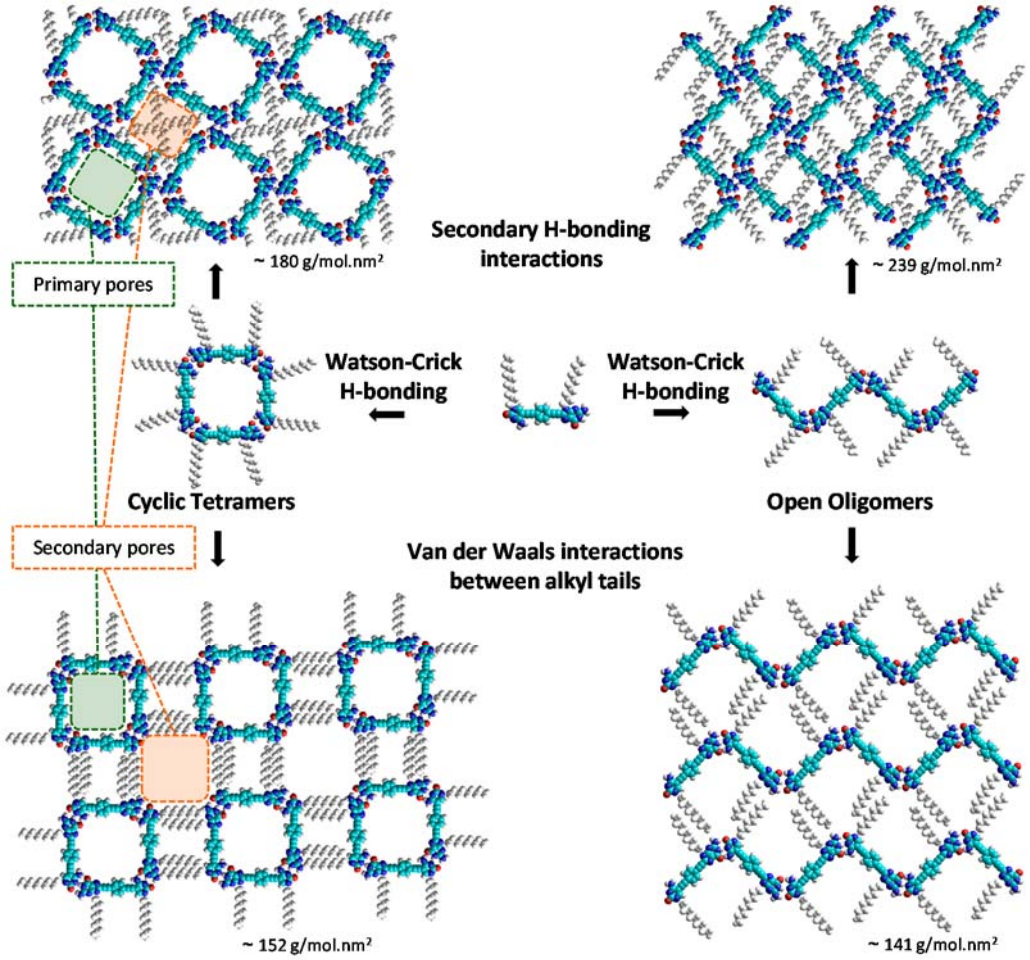


Figure 54. Proposed association modes for the **GC1** monomer depending on the main secondary non-covalent interactions, ruling cyclic tetramer/open polymer network formation. The density of each 2D network has been calculated.

Measurements.

GC1. After drop-casting a 5.0 x 10^{-6} M solution of the compound **GC1** in a 1:1 mixture of TCB:octanoic acid (OA) on top of the HOPG substrate, large organized domains of non-cyclic oligomers were observed (Figure 55).

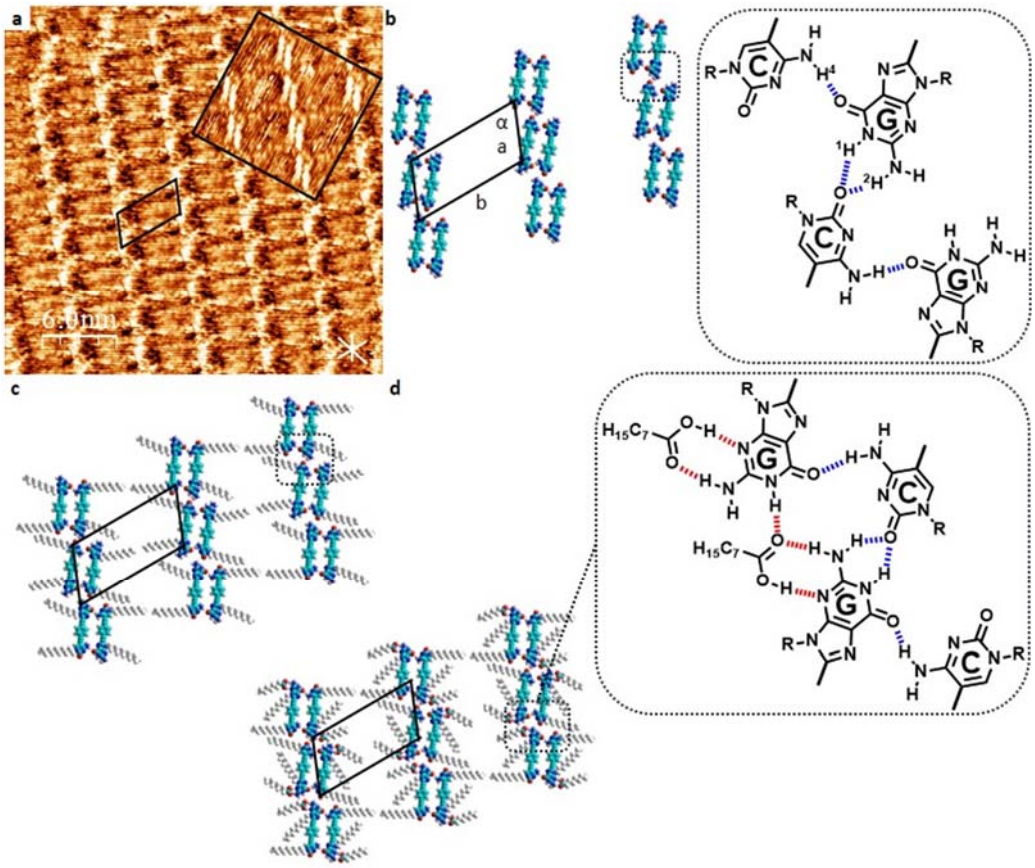


Figure 55. **Proposed association mode for GC1.** (**a**) STM image of **GC1** at the TCB:OA/HOPG interface (5.0 x 10^{-6} M; I_{set} = 50 pA, V_{bias} = -350 mV). The inset, showing the arrangement in monomer pairs, corresponds to another STM image of **GC1** on HOPG from a solution 6.0 x 10^{-6} M in TCB:OA 1:1 (I_{set} = 50 pA, V_{bias} = -350 mV). White lines indicate the normal axes of graphite. (**b**) Model along with scheme of the stabilizing motif. The unit cell is indicated in black lines (a = 2.7 ±0.1 nm, b = 4.5 ±0.1 nm, α = 68 ±1°). (**c**) Molecular model indicating the monomer decyl chains. (**d**) Model for the self-assembled network including OA solvent molecules along with scheme of the stabilizing motif.

The network did not fit any of the proposed patterns in Figure 54. Instead of establishing the expected Watson-Crick interactions, the model suggested that the ditopic π-conjugated monomers were arranged in parallel rows comprising packed pairs of molecules[216] that were associated *via* H-bonding interactions between the G-carbonyl group and the amino C-H^4 proton (Figure 55b). At the same time, the rows grew by additional H-bonding between the C-carbonyl group and the G-H^1 and G-H^2 protons. Although not clearly observed in the images, such double-chain pattern can leave all the peripheral long alkyl tails lying flat on the surface and arranged orthogonally with respect

[216] Similar 2D arrangements were observed from other structurally related ditopic monomers. See: A. Llanes-Pallas, M. Matena, T. Jung, M. Prato, M. Stöhr, D. Bonifazi, *Angew. Chem. Int. Ed.* **2008**, *47*, 7726–7730.

to the rows, so that they can establish multiple van der Waals interactions to stabilize the network (Figure 55c). Still, the distance between rows is too large to be explained solely by interdigitation. We believe that OA solvent molecules might also participate in the network establishing H-bonding interactions with the G-nucleobases and van der Waals interactions with the alkyl tails between rows (Figure 55d). **GC1** would be therefore reorganized onto HOPG to yield a densely-packed network, calculated as 228.0 g/mol·nm^2, that maximized molecule-molecule, molecule-substrate and, possibly, molecule-solvent interactions. Cyclic **GC1** assemblies are probably not preserved on the surface because they leave large void spaces, 2.2 nm^2 in diameter, that would lead to networks with much lower density, estimated as 180.3 g/mol·nm^2.

Hence, we reasoned that partial filling of the cavity with alkyl tails, introduced at the central benzene block (**GC2**), might be a strategy to increase surface density and favor the formation of cyclic tetramer networks.

GC2. When applying a drop of a **GC2** solution in OA onto the freshly cleaved HOPG surface, large domains of cyclic tetrameric species were now obtained (Figure 56). Watson-Crick molecular recognition between the two inserted nucleobases allowed the specific formation of cyclic species formed from four monomeric subunits. These discrete tetramers then assembled into the desired ordered nanostructured porous surface by stablishing H-bonding interactions between the G-aminopyridine fragments.

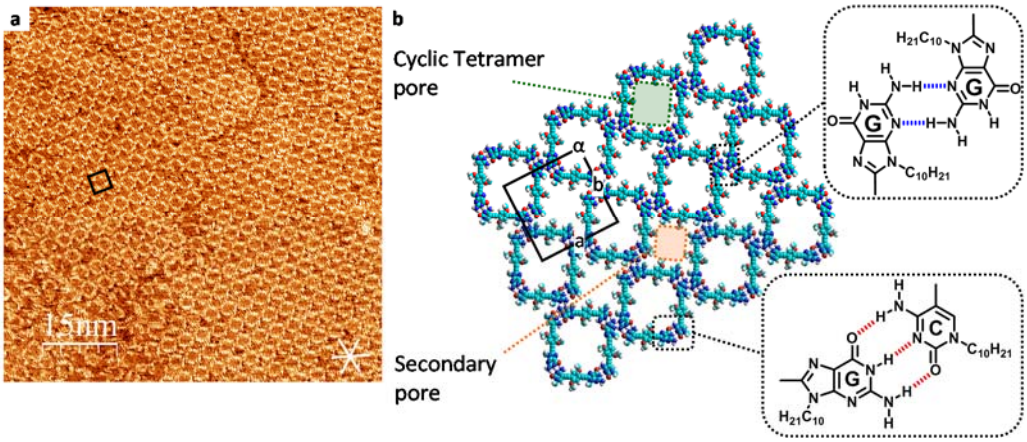


Figure 56 GC2 self-assembly into cyclic tetramers. (**a**) STM image of **GC2** on HOPG from a solution 7.1 x 10^{-6} M in OA (tunneling parameters: I_{set} = 280 pA, V_{bias} = -180 mV) along with (**b**) proposed model and H-bonded patterns schemes. Dashed red lines indicate Watson-Crick base-pairing and the dashed blue lines correspond to secondary H-bonding between nucleobases. The unit cell is indicated in black lines (a = 3.4 ±0.1 nm, b = 3.6 ±0.1 nm, α = 92 ±2 °). White lines indicate the normal axes of graphite.

Since **GC2** alkyl chains were not observed in the recorded images, they were not represented in the proposed model (Figure 56b). Monomer **GC2** is provided with two long –C_{10} chains at the bases and one –OC_8 chain in the block that do not fit in the space left

between four cyclic tetramers, according to the Hyperchem model. It is assumed that the central $–OC_8$ tails are grouped within the cyclic tetramer cavity, increasing surface density, but most of the peripheral alkyl tails must be backfolded in the supernatant solution. Some distortion in the tetramers is possible: the unit cell parameter along one axis can be shorter than the distance in the other axis. This is often due to the difference in alkyl chain coadsorption: more in one direction, less in the other, and the extra chains desorbed in solution.

It is interesting to note that despite the great affinity of long alkyl chains for the HOPG substrate, these were not physisorbed as a whole. They could be adsorbed on it and the network of cyclic tetramers could be stabilized through van der Waals interactions *via* their interdigitation (as shown in Figure 54). Desorption of the alkyl chains results in an unstable situation but it is compensated by the establishment of H-bonds between peripheral groups in the nucleobases, and concretely between G-aminopyridine fragments. Therefore, non-covalent H-bonding interactions clearly have a dominant effect in driving the network self-assembly, overcoming van der Waals interactions. In this way, a more densely-packed network occured. As a matter of fact, despite carrying various long alkyl chains, **GC2** displays an extremely low solubility in almost any solvent and mixtures of solvents, whereas the bulky ribose analogue $\mathbf{GC_{rib1}}$ (see the Background and Objectives section) is perfectly soluble. This difference in solubility cannot be solely explained by interaction between alkyl chains or by π-π stacking. Secondary H-bonding interactions between aminopyridine fragments, leading to further aggregation are probably responsible for the insolubility of **GC2**. The addition of OA turned out to be very useful for obtaining a clear stock solution. OA may limit the aggregation by competing with the G-nucleobases in stablishing double H-bonding with the aminopyridine fragments (see Figure 55d).

In view of these results, we decided to synthesize a third-generation monomer (**GC3**), which is now equipped with shorter $–C_2$ and $–C_6$ chains at the G and C bases, respectively. The length of these alkyl groups was precisely tailored at each base in order to completely fill the secondary unspecific pores generated between 4 cyclic tetramers and, at the same time, to allow the chains to physisorb onto HOPG, thus promoting additional interactions to stabilize the network.

GC3. The most stable network on the surface was reached after applying a 8.2×10^{-6} M solution of **GC3** to the HOPG substrate. Indeed, large domains of cyclic tetrameric species were obtained (Figure 57).

Watson-Crick molecular recognition between the nucleobases allowed the specific formation of the cyclic tetramers. Each cyclic tetramer was formed by four monomeric subunits and these were brought together through H-bonding between the G-aminopyridine fragments. The tetrameric networks showed remarkably high stability at the HOPG/OA:TCB interface, when the molecules were concentrated on the surface. Excellent coverage of the surface and a higher resolution were a result of this network stability: **GC3**

molecules formed large single domains whose size extended well beyond 100 x 100 nm^2 (Figure 57a). The inner pore of the cyclic tetrameric species was filled by the octyloxy chains of the central blocks as observed in higher resolution images (Figure 57c), whereas some contrast in the shape of four lobes appeared in the cavity between tetramers, which indicated physisorption of the cytosine $–C_6$ chains.

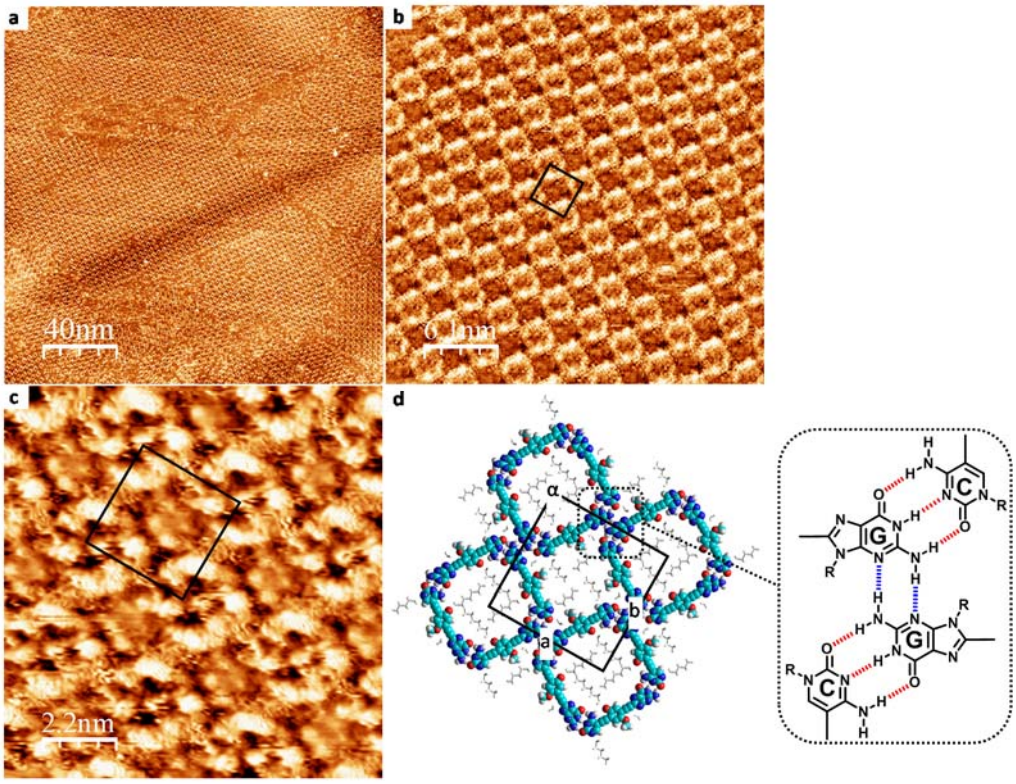


Figure 57 Self-assembled porous network formed by GC3 on HOPG from a solution 8.2 x 10^{-6} M in TCB:OA (1:1) (tunneling parameters: I_{set} = 200 pA, V_{bias} = -300 mV). (**a**) Large scale STM image. (**b**) Smaller scale image. (**c**) High-Resolution STM image. (**d**) Proposed model for **GC3**, along with scheme of the stabilizing motif. Dashed red lines indicate Watson-Crick base-pairing and the dashed blue lines correspond to secondary H-bonding between nucleobases. The unit cell is indicated by black lines (a = 3.6 ±0.1 nm, b = 3.6 ±0.1 nm, α = 89 ±1 °).

A growth of the domains with time, driven by thermodynamics, was also observed (Figure 58). "Ostwald ripening" is a well-documented phenomenon at the solid-liquid interface and corresponds to the observation that large domains grow at the expense of smaller domains, resulting in an increase of the degree of ordering.[217] The thermodynamic driving force is the reduction of the circumference-to-area ratio and thereby the lowering of the interfacial, or line, energy. At a domain boundary, molecules are often not ideally

[217] A. Stabel, R. Heinz, F. C. De Schryver, J. P. Rabe, *J. Phys. Chem.* **1995**, *99*, 505–507.

close-packed and the free volume within the 2D polycrystal is significantly increased. Therefore, individual molecules or lamella fragments at the domain borderline can change their orientation with respect to the underlying substrate without diffusing within the single crystals and without transition to the supernatant solution. At the solid-liquid interface, the adlayer is in equilibrium with the supernatant solution and there is a dynamic exchange of molecules. These exchange processes are important for the formation of supramolecular structures: the exchange of molecules in two and in three dimensions (monolayer-solution) provides a "healing" mechanism to repair defects.[218]

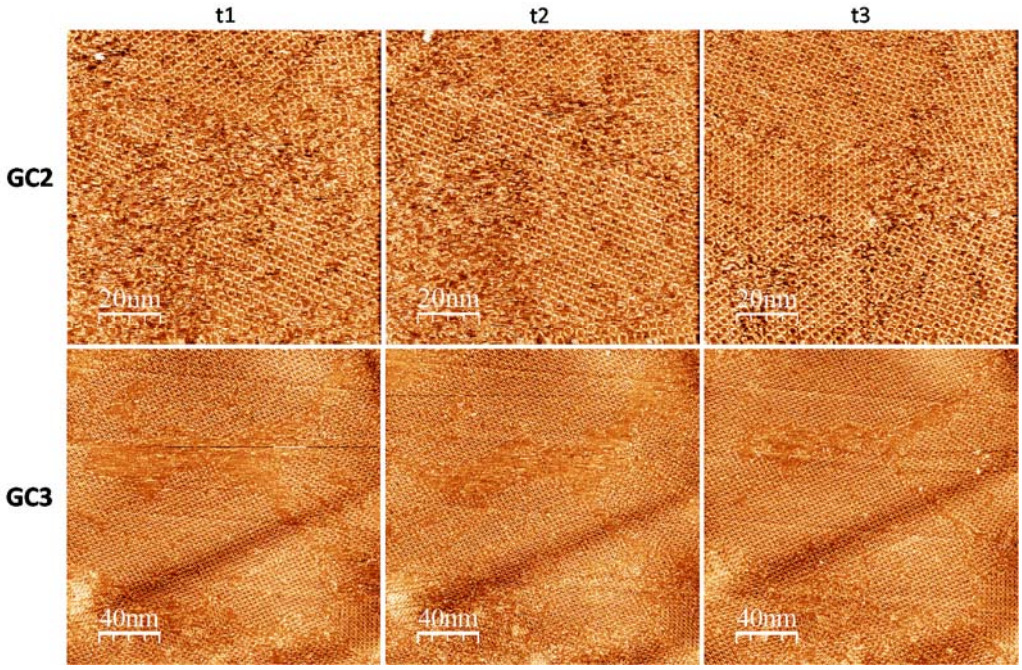


Figure 58. Large scale STM images taken over 45 min of GC2 (7.1 x 10^{-6} M; I_{set} = 300 pA, V_{bias} = -170 mV) and **GC3** (8.2 x 10^{-6} M; I_{set} = 200 pA, V_{bias} = -300 mV) showing how the domains grow toward full surface coverage.

In summary, a highly dense non-cyclic organization was formed by **GC1**, where all the long decyl chains were adsorbed on the surface. Cyclic networks were promoted by partial filling of the pores by octyloxy chains. As a result, the physisorption on HOPG of **GC2** allowed the construction of a well-structured 2D system composed by H-bonded cyclic tetrameric species. **GC3**, equipped with shorter $-C_2$ and $-C_6$ chains, afforded an even more stable dense self-assembly where the molecule-substrate interactions were maximized. At this point, we considered the preparation of a fourth monomer (**GC4**). This one has, on one hand, an unsubstituted benzene ring as central block; and on the other, carries short ethyl

[218] S. De Feyter, A. Gesquière, M. M. Abdel-Mottaleb, P. C. M. Grim, F. C. De Schryver, C. Meiners, M. Sieffert, S. Valiyaveettil, K. Müllen, *Acc. Chem. Res.* **2000**, *33*, 520–531.

and hexyl chains at the G and C bases, respectively (like **GC3**). Our aim was to destabilize the non-cyclic network formed by **GC1** by incorporating alkyl chains that broke such an assembly, but did not take up the total space and left empty cavities. The proposed alkyl substitution fitted perfectly in the space left between four cyclic tetramers and the alkyl tails could be physisorbed onto HOPG, as was seen for **GC3**, thus stabilizing a dense porous network where the inner cavity would be totally empty.

GC4. Monomer **GC4** will be studied by STM at the solid-liquid interface in the near future with the aim to obtain cyclic networks with totally empty inner cavities, and not occupied with alkoxy chains.

2.2.2. Host-Guest Systems.

Large domains of stable porous networks have been obtained. However the cavity of the cyclic tetramer has been partially sacrificed in the process by the incorporation of octyloxy chains. At this point, we would like to see if this pore can host a guest molecule. Coronene (**cor**) was chosen as the guest molecule for these studies because it fitted well in the pore of the cyclic tetramer formed by **GC3**, together with its commercial availability. Also, **cor** is a polycyclic aromatic hydrocarbon, composed of six peri-fused benzene rings (Figure 59a), shows strong affinity for the HOPG substrate and appears as a bright disk in the STM images, originating a conveniently recognizable feature. However, single domains of **cor** cannot be imaged because of poor molecule-molecule interactions. Still, **cor** molecules have a strong tendency to incorporate themselves in molecular networks on the surface, making them suitable guests for the study of the host ability of our systems. Phthalocyanines were also considered as potential guest molecules for these host-guest studies. Cu^{II}-phthalocyanine (**CuPc**) is a flat aromatic molecule able to strongly adsorb onto HOPG. Moreover, **CuPc** possess a cross-like shape, well-suited for these squared host cavities (Figure 59b). By adding a guest, a competition to fill the pore between the octyloxy chains in the central block of the molecule (when present) and the guest molecule must be stablished.

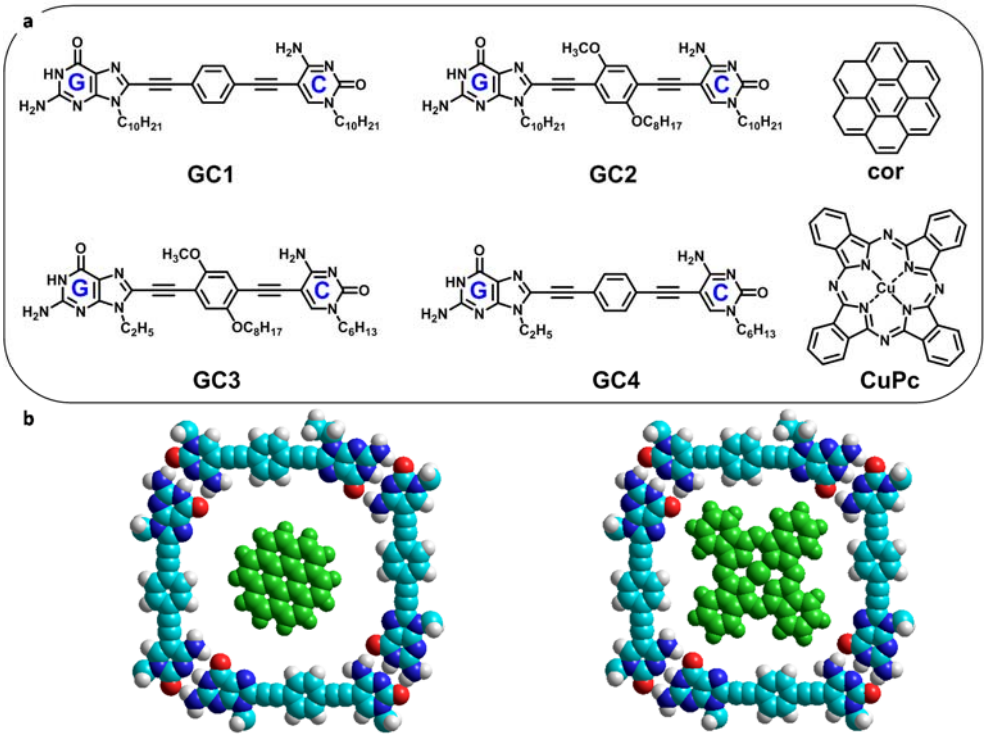


Figure 59 Host-guest systems. (**a**) Chemical structure of the candidates for the host-guest studies and (**b**) general models of their potential host-guest systems.

GC3+cor. To investigate the anticipated ability of the created nanoporous surfaces to host an appropriate molecule, a premixed solution of the **GC3** monomer with an excess of **cor** was prepared. When applying this mixture onto HOPG, medium-sized domains covering large areas were observed (Figure 60d-g). The bright feature inside the cavities was a strong evidence for the co-adsorption of **cor**. Also, when changing the sign of the bias voltage, interesting well-resolved inverted polarity images were obtained (Figure 60d). These images must be interpreted as being opposed to the others. That is, bright features corresponded to low electron density regions, such as empty spaces. In Figure 60a, the empty cyclic tetramers are seen as dark squares surrounded by inner and outer regions of similar brightness. However, when **cor** was present, a higher contrast could be appreciated: the inside of the tetramer was not so bright anymore, giving evidence of the adsorption of the guest molecule. The octyloxy chains were no longer observed and were, very likely, backfolded in the supernatant solution, leaving room to the guest molecule. This means that an equilibrium, which was shifted to the adsorption of the guest molecule by the pore, was taking place when both molecules were adsorbed onto HOPG from solution. In the domain borders, fuzzy areas were observed, which were evidence of this adsorption/desorption equilibrium between the octyloxy chains and **cor** (Figure 60h). As the exchange was faster than the scanning, it was seen as an average image. However,

when adding **cor** to the liquid phase after the formation of the network of cyclic tetramers in a sequential manner, no host-guest system was obtained, even when the guest molecule was added in excess. In other words, in a sequential addition, the $-OC_8H_{17}$ chains did not desorb and **cor** molecules were not imaged.

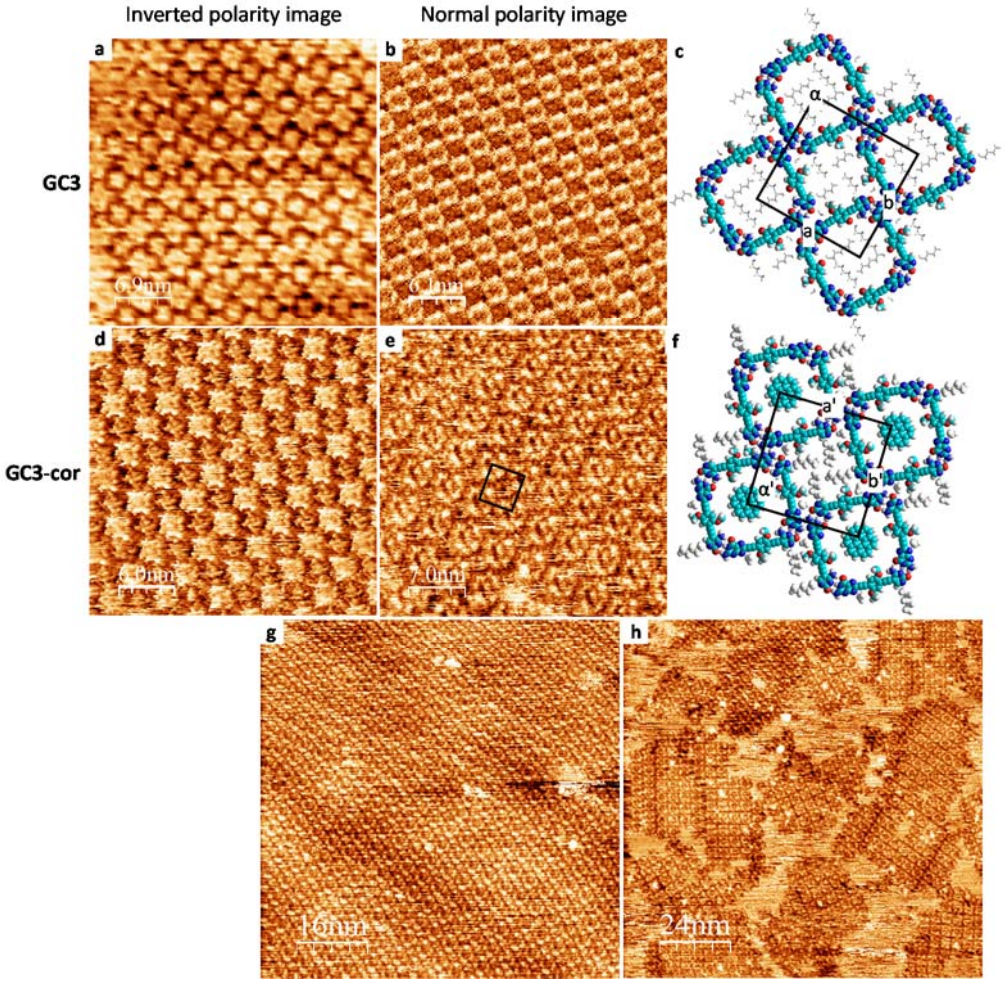


Figure 60. Self-assembly of the GC3 one-component and GC3-cor systems at the HOPG/TCB:OA (1:1) interface. Normal polarity tunneling parameters: I_{set} = 80 pA, V_{bias} = -350 mV. Inverted polarity tunneling parameters: I_{set} = 80 pA, V_{bias} = 350 mV. (**a**) Inverted polarity STM image of a single **GC3** domain. (**b**) Normal polarity STM image of **GC3**. (**c**) Model proposal for the one-component system. The unit cell is indicated by black lines (a = 3.6 ±0.1 nm, b = 3.6 ±0.1 nm, α = 89 ±1°). (**d**) Inverted polarity STM image of **GC3** and **cor** (1:200; 6.0 x 10^{-6} M). (**e**) Small scale normal polarity STM image of the **GC3**-**cor** system (**f**) Model proposal for the host-guest system (a' = 3.6 ±0.1 nm, b' = 3.5 ±0.2 nm, α' = 89 ±1°). (**g**) Large normal polarity scale STM image of the **GC3**-**cor** system. (**h**) Normal polarity STM image where bright fuzzy areas between domains are observed, indicating guest exchange.

GC2+cor. H-bonding strength between nucleobases in comparison with van der Waals interactions was again proven when making the deposition of a premixed solution of monomer **GC2** and **cor**. A similar organization of the 2D nanoporous surface to the host-guest system formed by **GC3** and **cor** was imaged (Figure 61a). It is noteworthy that in this **GC2**-**cor** system only few alkyl chains can be adsorbed on the HOPG surface. We estimated that two $-C_{10}$ and one $-OC_8$ chains *per* molecule were desorbed and thus negatively contributing to the stabilization of the network. Again, a situation where multiple secondary H-bonding interactions intervene was preferred to a less densely-packed assembly, like the one proposed in Figure 54, where numerous van der Waals molecule-molecule and molecule-surface interactions could positively contribute.

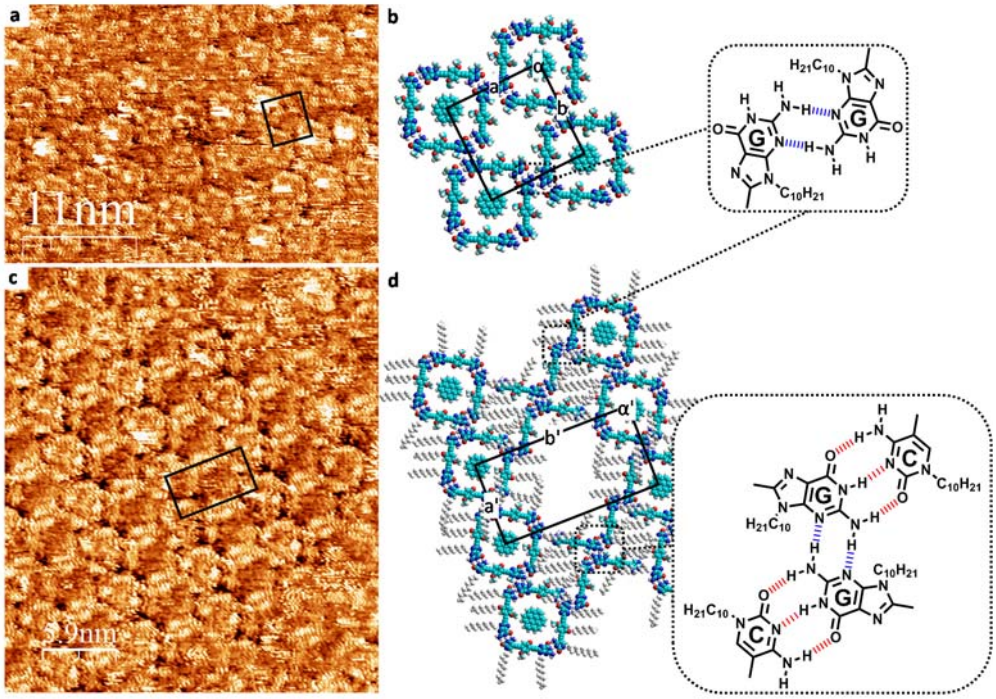


Figure 61. GC2-cor self-assembly at the HOPG/OA interface (tunneling parameters: I_{set} = 250 pA, V_{bias} = -170 mV). (**a**) STM image corresponding to domain pattern 1 and (**b**) molecular model. The unit cell is indicated by black lines (a = 3.3 ±0.1 nm, b = 2.6 ±0.2 nm, α = 89 ±1 °). (**c**) STM image for domain pattern 2 (a' = 6.5 ±0.1 nm, b' = 2.5 ±0.2 nm, α' = 89 ±3°) and (**d**) corresponding proposed model along with schemes of the H-bonded motifs.

In comparison with the case of **GC3**, smaller domains were observed for the **GC2**-**cor** system, where **cor** molecules were hosted in the cyclic tetramers, proving that the system was less stable. Additional small domains of another pattern were imaged (Figure 61c). The proposed model for this second arrangement corresponded to an uncompleted picture in the formation of the network of cyclic tetramers bound through secondary H-bonding interactions between aminopyridine fragments. In one direction, rows of cyclic tetramers

bound through their G-aminopyridine external fragments were observed. However, in the other direction of the plane, an incomplete row of tetramers was intercalated between two assembled rows of macrocycles. Although not appreciated in the images, in this uncompleted arrangement, all the long alkyl chains of the monomers, as well as **cor** molecules, could be potentially adsorbed onto HOPG.

GC1+cor. Despite the fact that **GC1** did not form a nanoporous network composed of cyclic tetramers, the deposition on HOPG of the system with **cor** was essayed to see if the guest molecule could template the cyclization process. In several cases, guest molecules have been found to induce structural changes into the self-assembling system, only by their presence.[219] Well-resolved STM images of large domains formed by a less regular arrangement than the previous ones (without **cor**) were obtained (Figure 62).

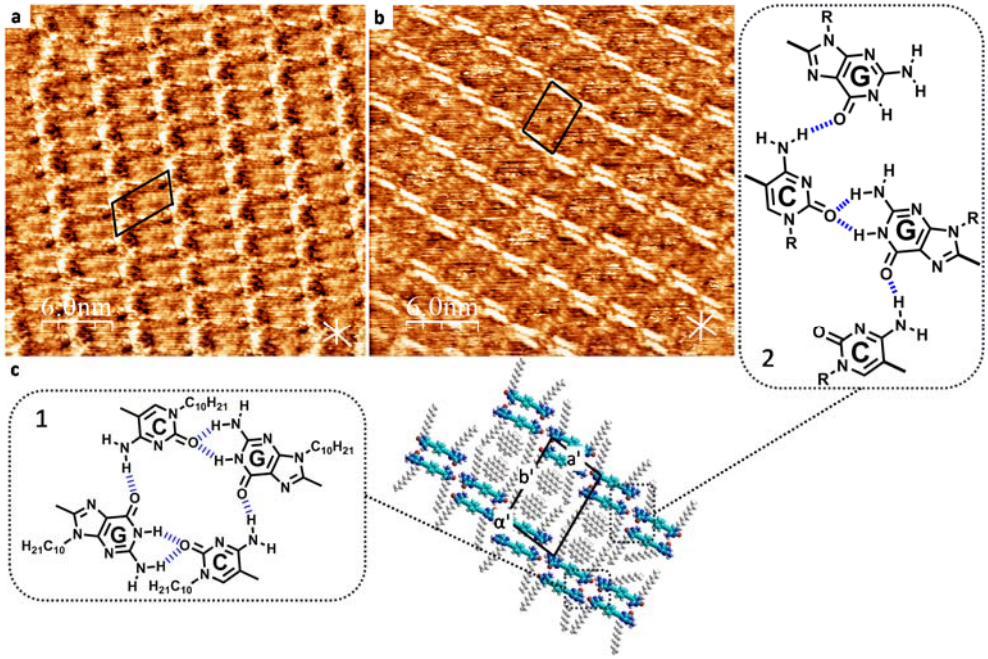


Figure 62. GC1 one-component and GC1-cor systems self-assemblies. (**a**) STM image of **GC1** at the TCB:OA/HOPG interface (5.0 x 10^{-6} M; I_{set} = 50 pA, V_{bias} = -350 mV). The unit cell is indicated by black lines (a = 2.7 ±0.1 nm, b = 4.0 ±0.1 nm, α = 87 ±1 °). (**b**) STM image of **GC1-cor** at the HOPG-TCB/OA (1:1) interface (6.0 x 10^{-6} M; I_{set} = 50 pA, V_{bias} = -300 mV). White lines indicate the normal axes of graphite. (**c**) Corresponding proposed molecular model where **cor** molecules have been tentatively incorporated, along with schemes of the recurring H-bonded motifs (a' = 3.0 ±0.2 nm, b' = 4.0 ±0.2 nm, α' = 94 ±3°).

[219] a) S. Furukawa, K. Tahara, F. C. De Schryver, M. Van der Auweraer, Y. Tobe, S. De Feyter, *Angew. Chem. Int. Ed.* **2007**, *46*, 2831–2834; b) M. Blunt, X. Lin, M. d. C. Giménez-López, M. Schröder, N. R. Champness, P. H. Beton, *Chem. Commun.* **2008**, 2304–2306; c) D. Wu, K. Deng, M. He, Q. Zeng, C. Wang, *Chemphyschem* **2007**, *8*, 1519–1523; d) M. Li, P. Xie, K. Deng, Y.-L. Yang, S.-B. Lei, Z.-Q. Wei, Q.-D. Zeng, C. Wang, *Phys. Chem. Chem. Phys.* **2014**, *16*, 8778–8782; e) D. Bléger, D. Kreher, F. Mathevet, A.-J. Attias, G. Schull, A. Huard, L. Douillard, C. Fiorini-Debuischert, F. Charra, *Angew. Chem. Int. Ed.* **2007**, *46*, 7404–7407.

The STM image showed a very similar pattern to the one obtained without **cor**: characteristic bright double rods that elongated in one "*x*" direction of the plane following a stair-like decay. From the top-ends of the bright sticks, two darker dashed lines were disposed in the "*y*" direction, probably corresponding to $-C_{10}$ alkyl chains. We proposed an association mode, close to the one presented by monomer **GC1** on its own (Figure 55), where **cor** molecules may be randomly incorporated in the self-assembled web. Some dim bright disk-shaped forms can be appreciated in some of the spaces left between double stripes that could hypothetically correspond to guest molecules. We believe that the system was mainly stabilized through van der Waals interactions between alkyl tails and **cor** molecules. Up to two guest molecules *per* **GC1** dimer can be encapsulated in the flexible open-oligomer-like network formed solely by **GC1**. The STM images showed that the obtained network was not ideally regular, the two-molecule system appeared to be a bit loose, causing some translational defects (Figure 62c, compare schematic H-bonding motifs 1 and 2). Even though the self-assembled architecture would be mainly stabilized through van der Waals interactions, some stabilizing H-bonding interactions between monomers seemed to be intervening. Taking into account that our species present a large number of available sites for H-bonding, numerous stabilizing interactions are possible between them. In particular, the H-bonded pattern formed between two G-nucleobases and OA, described in Figure 55d for the one-component **GC1** system, could be occurring here as well.

The system formed by **GC1** and **cor** presented STM images with higher resolution than **GC1**, maybe because the **GC1**-**cor** system was more compact and the mobility of the molecules decreased. However, smaller domains and some irregularity in the packing were observed.

GC1-GC3+CuPc. Different premixed solutions of **CuPc** with molecules of the **GC1**-**GC3** family, covering a large range of concentrations (3.0×10^{-6} M – 1.0×10^{-5} M) were prepared and dropcasted onto HOPG. However none of them afforded a host-guest system, even when adding **CuPc** in excess. Sequential addition was ineffective as well. Alike **cor**, **CuPc** failed to template the cyclization process of **GC1**, despite the fact that **CuPc** was endowed with a larger aromatic surface and fitted better the pore of the tetramer in terms of size and shape. In the case of monomers **GC2** and **GC3**, the *p*-disubstitution of their central block with methoxy and octyloxy chains will decrease the area of the cavity and may impede the inclusion of **CuPc**, even if the chains desorb.

GC4+cor, GC4+CuPc. Molecule **GC4** has been prepared with the aim of creating a nanoporous network of self-assembled macrocycles where the inner cavity of the tetramers is totally empty. Therefore, stable host-guest networks are expected both with **cor** and **CuPc** guest molecules. Let's remark that in this case, the two deposition techniques (premixed solution and sequential addition) might work, since no adsorption/desorption equilibrium between the guest and a fragment of the monomer should occur. Additionally, as previously stated, **CuPc** might afford a very stable **GC4**-**CuPc** species since this guest fits

perfectly in the cavity of this tetramer in terms of shape and symmetry. This experiments at the solid-liquid interface will be soon carried out.

2.2.3. Surface Chirality.

As mentioned in the Introduction (section **2.4.**), the formation of 2D chiral organizations from nonchiral molecules, where chirality is generated and directed by the arrangement of molecular building units in the 2D assemblies, is an intriguing phenomenon.[220] In the case of our cyclic tetramer system, different pore chirality was observed in different domains (Figure 63). This chirality on the surface was given by the side of the monomer in contact with the surface. Assuming an arbitrary cycle sense taking the H-bonding interaction from G to C, we could define a clockwise (CW) and a counter-clockwise (CCW) tetramer on the surface. Let's note that different domains with different pore chiralities were also recognized in the **GC2** networks.

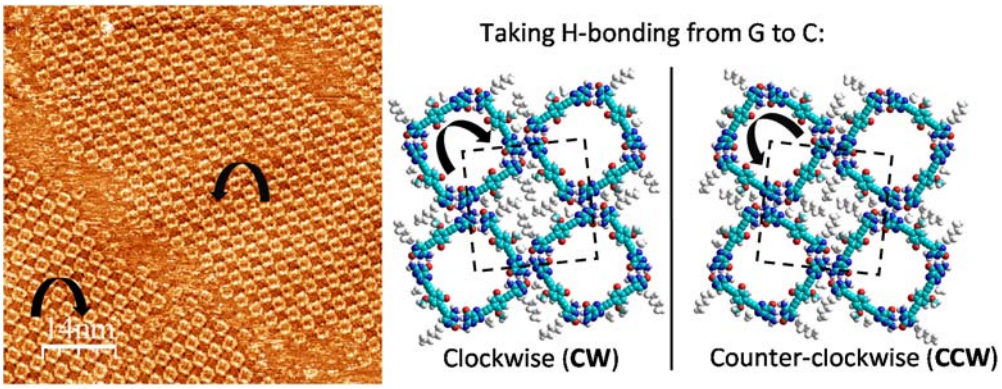


Figure 63. STM image of **GC3** on HOPG from a solution 8.2 x 10^{-6} M in TCB:OA (1:1) (tunneling parameters: I_{set} = 200 pA, V_{bias} = -300 mV) showing two domains with different pore chirality and explanatory model.

Similarly to racemic mixtures of chiral molecules, this was a case of spontaneous resolution of a prochiral supramolecular system. We had in our hands an example of a 2D crystal conglomerate with chiral domains and chiral unit cells. They formed on the surface, as the association between tetramers of the same stereochemistry was more favorable than between opposite enantiomers. In the area between both chiral domains, unordered regions were observed, where the molecules were probably in rapid exchange.

[220] a) *Supramolecular Chirality* (Eds.: M. Crego-Calama, D. N. Reinhoudt), Springer-Verlag, Berlin, Heidelberg, **2006**, *265*; b) L. G. Teugels, L. G. Avila-Bront, S. J. Sibener, *J. Phys. Chem. C* **2011**, *115*, 2826–2834; c) J. Elemans, I. De Cat, H. Xu, S. De Feyter, *Chem. Soc. Rev.* **2009**, *38*, 722–736; d) J. Liu, T. Chen, X. Deng, D. Wang, J. Pei, L. -J. Wan, *J. Am. Chem. Soc.* **2011**, *133*, 21010–21015.

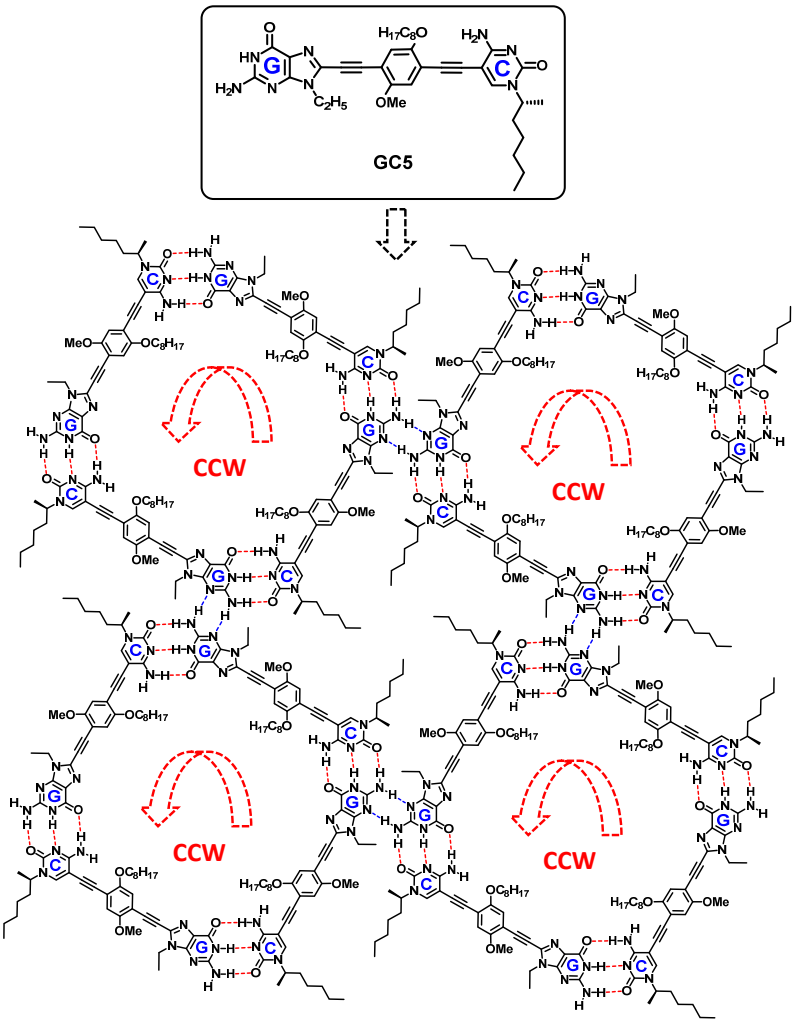


Figure 64. Synthesized candidate for single pore chirality on the surface.

At this point, the self-assembly of a stereochemically pure monolayer was addressed. In order to explore if only one kind of pore (CW or CCW) and a single chiral domain could be obtained, an enantiomerically pure monomer (**GC5**) was designed (Figure 64). This molecule was based on the structure of **GC3** and should form stable networks of cyclic tetramers. In the **GC3** STM images, the hexyl chain of the C base seemed to be totally adsorbed on the surface, and for this reason, we expected that the incorporation of an asymmetric carbon atom, bearing an additional methyl group within this chain, could provoke some steric hindrance in the interaction between substrate and molecule. Consequently, the $-C_6$ chain could be disposed so that the methyl substituent points upwards, for instance. This disposition would involve that always the same face of the monomer would be adsorbed onto HOPG and therefore, a preference toward CW or CCW

pores should be appreciated. Moreover, if a predilection for only one type of pore were to occur, this could have an effect in the overall organization. Actually, the packing between different CW and CCW domains is uneven and hampers the growth of the domain. A single domain may, thus, lead to larger and better ordered 2D networks.

On these grounds, monomer **GC5** was synthesized and will be soon studied by STM at the solid-liquid interface. If tetramer and domain's chirality is controlled, "sergeant and soldiers" studies,[221] could be contemplated in order to investigate amplification of chirality phenomena. In those experiments, we would analyze the amount of the homochiral material (**GC5**), in a mixture with a large amount of an achiral monomer (**GC3**), needed to trigger the formation of a supramolecular assembly endowed with the chirality preferred by **GC5**. Additionally, after synthesizing the monomer with opposed chirality, "majority rules" studies could also be carried out, where a slight majority of one of the enantiomers could produce a disproportionately large preference for one pore sense domain.[222]

2.2.4. Pore Size Control and Size-selective Guest Adsorption.

Our supramolecular system affords the possibility to form cyclic species with different sizes by simply changing the length of the central block taken from our molecular toolkit. In that simple manner, a display of well-ordered 2D networks with a wide range of cavity sizes could be obtained. One key objective is to obtain nanostructured networks with tunable pore size that could serve as hosts for size-matching guests that would selectively fill the pores. Recognition processes at the solid-liquid interface are fairly complex, since avoiding phase separation of molecular components is a major challenge.[223] As it has been exposed earlier, the idea of immobilizing different functional units in a finely-tuned 2D porous network can provide interesting applications in emerging fields such as molecular-scale electronics, catalysis or sensing.

We focused again on strongly bound GC monomers (Figure 65a). We targeted two short molecules (**GC6** and **GC7**) able to host small guest molecules, like phenalene, in the self-assembled network of cyclic tetramers (Figure 65b, left). According to our molecular models, a guest molecule like triphenylene would probably fit too tightly in the pore cavity, but might be included since our self-assembled system can adapt to some extent. Smaller guests like phenalene or C_{60}, which only require a small area to adsorb on surfaces due to its 3D structure, would probably be better size-matching guests. A medium-sized monomer like **GC4** may host a guest molecule like **CuPc** (Figure 65b, center), providing we are able to generate a fully empty cavity (*vide supra*). A final monomer (**GC8**) was designed, long

[221] L. J. Prins, P. Timmerman, D. N. Reinhoudt, *J. Am. Chem. Soc.* **2001**, *123*, 10153–10163.
[222] J. van Gestel, A. R. A. Palmans, B. Titulaer, J. A. J. M. Vekemans, E. W. Meijer, *J. Am. Chem. Soc.* **2005**, *127*, 5490–5494.
[223] a) Y. Xue, M. B. Zimmt, *Chem. Commun.* **2011**, *47*, 8832–8834; b) K. E. Plass, K. M. Engle, K. A. Cychosz, A. J. Matzger, *Nano Lett.* **2006**, *6*, 1178–1183; c) K. S. Mali, B. Van Averbeke, T. Bhinde, A. Y. Brewer, T. Arnold, R. Lazzaroni, S. M. Clarke, S. De Feyter, *ACS Nano* **2011**, *5*, 9122–9137; d) Y. Xue, M. B. Zimmt, *J. Am. Chem. Soc.* **2012**, *134*, 4513–4516.

enough to induce a considerable big pore. The cyclic tetramer formed by **GC8** should be able to incorporate large guests, like naphthalocyanines (Figure 65b, right), having as well interesting electronic properties.[224]

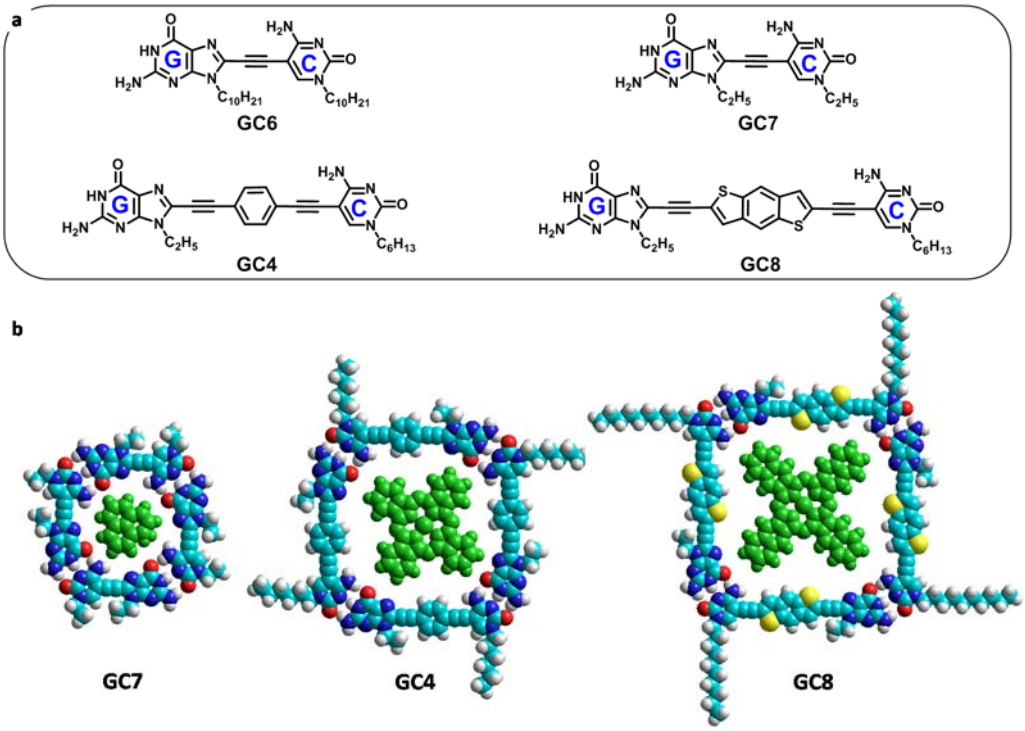


Figure 65. Pore-size control. (**a**) Chemical structure of the candidates for size discrimination studies and (**b**) models of their potential host-guest systems.

In molecules **GC6** and **GC7**, the G and C moieties were directly connected *via* a triple bond. **GC6** was equipped with long $–C_{10}$ alkyl chains, whereas **GC7** carried shorter $–C_2$ tails. Both could potentially form cyclic species with the same pore diameter, able to host equivalent guest molecules in terms of size. We foresaw that **GC6**, having longer alkyl chains, could undergo a different 2D porous network, stabilized through van der Waals interactions, not as densely packed as the one engineered by **GC7**. We expected that candidate **GC4** could host, besides **cor**, a slightly bigger guest like **CuPc** and give more stable host-guest systems since no alkyl chain would have to desorb from the HOPG surface. **GC8** contained as central block a benzodithiophene residue, longer than previous phenyl derivatives. This molecular candidate may induce the formation of a network with cavities with a diameter of 5.6 nm^2. Let's point out that despite the bigger length of the monomer, the number of rotatable bonds stayed the same, since the central block is a rigid aromatic heterocycle.

[224] P. Liljeroth, J. Repp, G. Meyer, *Science* **2007**, *317*, 1203–1206.

First, the one-component system will be studied in order to check the ability of the monomer to self-assemble into the desired cyclic tetrameric species. Then, the host-guests systems will be considered.

Small-sized pores.

GC6. Successive well-resolved images of small- to medium-sized domains of **GC6** showed that the monomer self-assembled on the surface into cyclic tetramers (Figure 66). Then, H-bonding between aminopyridine groups afforded a row of tetramers. These rows elongated following one axis ('*x*') of the plane, but in the orthogonal direction ('*y*'), the rows were separated. Van der Waals interactions between alkyl tails were presumed to be responsible for further stabilization of the network. We believed that the co-solvent (octanoic acid) could also play an important role in the stabilization of this structure and alternate on the surface between $-C_{10}H_{21}$ chains (Figure 66c). Alkyl groups were often not resolved in the images, which could be attributed to their mobility. The successive scans showed that the domains grew; and sometimes, a condensation of the network could be seen: the rows of tetramers came closer together (see Figure 66b). When this happened, H-bonding between aminopyridine fragments took over van der Waals interactions, in the secondary stabilization of the network. The STM images exhibited a flexible structure, mainly in the direction '*y*' of the plane, probably due to the mobility of the alkyl chains and the fact that van der Waals interactions are less directional than H-bonding. This is why the network was not perfectly steady and there were some fluctuations in the '*b*' and '*α*' values of the unit cell. Let's remark that this network was more difficult to stabilize than **GC3** because of the smaller size of the cyclic tetramers, and hence, of the space between them, which provoked desorption of the chains (see Figure 66d).

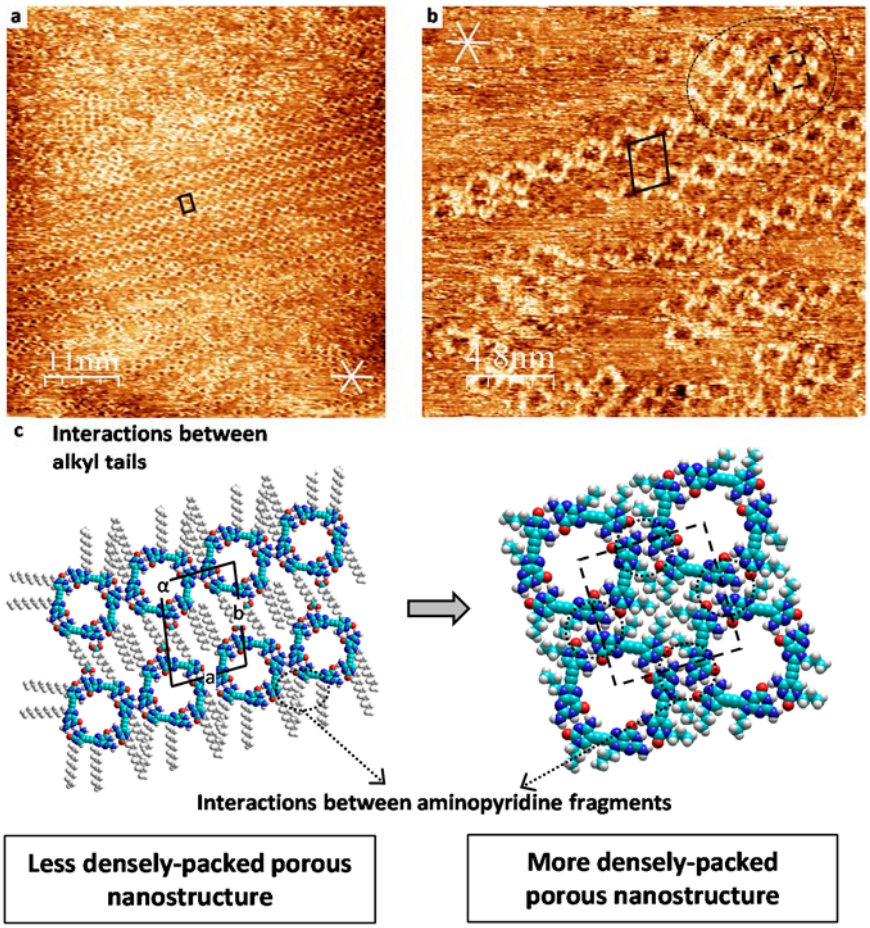


Figure 66. **Self-assembly of GC6 at the (1:1) TCB:OA/HOPG interface**. (**a**,**b**) STM images of a 4.1 x 10^{-6} M **GC6** solution (I_{set} = 20 pA, V_{bias} = -180 mV). The unit cell is indicated by black lines (a = 2.5 ±0.1 nm, b = 4.4 ±0.2 nm, α = 84 ±4°). White lines indicate the normal axes of graphite. (**c**) Proposed self-assembly behavior.

In view of these results, we decided to prepare and study monomer **GC7**. As explained before, this molecule is an analog of **GC6** but was functionalized with shorter alkyl chains that should not hinder the formation of a dense network of cyclic tetramers H-bonded through their aminopyridine fragments.

GC7. Molecule **GC7** afforded medium- to big-sized self-assembled domains of small-sized porous cyclic tetrameric species. A similar situation to molecule **GC6** was observed for **GC7**. Due to the lower steric interactions between tetramers, the system was, nevertheless, more stable on the surface and good images were more easily obtained. Octanoic acid may play again a crucial role in 2D ordering by stabilizing rows of cyclic tetrameric species *via* bonding competition with the aminopyridine fragments. It is interesting to note that OA molecules alone do not show a great tendency to form immobilized monolayers on

graphite. After all, this is one of the reasons why OA is used as solvent in STM. Nonetheless, OA and TCB solvent molecules could be trapped in the matrix formed by **GC7**, where a large area of HOPG was available for molecule adsorption between rows of tetramers. Co-adsorption of other small molecules at the organic liquid-solid interface and the solvent influence in the pattern observed has been also demonstrated earlier.[104,225] In our case, darker lines, probably corresponding to alkyl chains, stood out between macrocycles in the image (Figure 67b). Additionally, bright features localized in the corners of the tetramers that do not stablish secondary aminopyridine-aminopyridine interactions could be distinguished, presumably related to the carboxylic moiety bound to a G-fragment (Figure 67c).

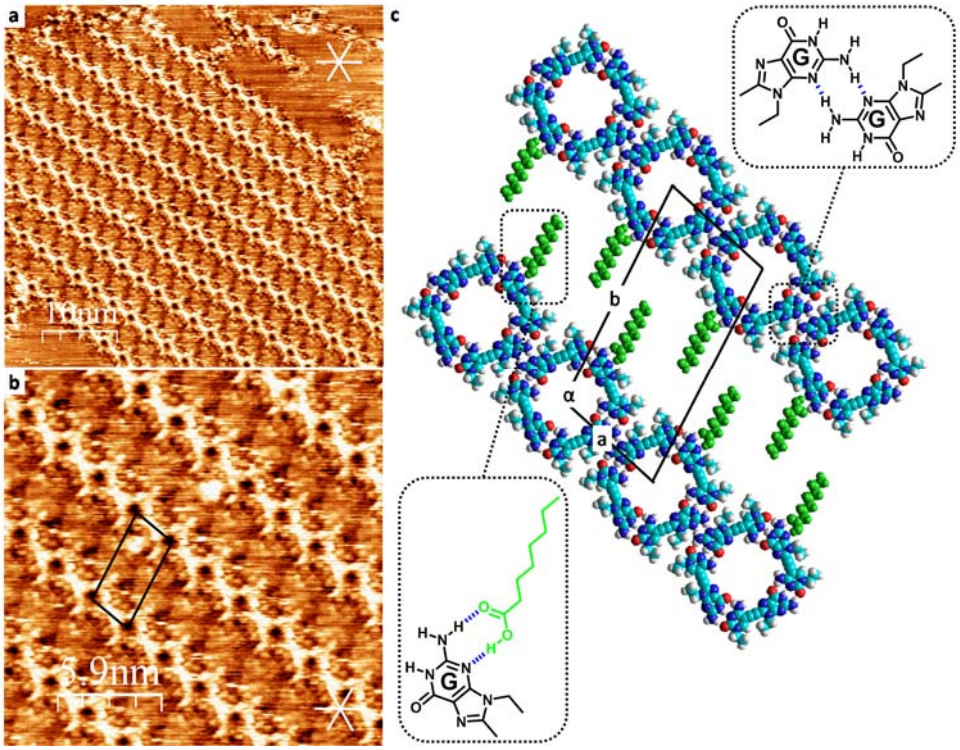


Figure 67. GC7 self-assembly at the HOPG/TCB:OA (1:1) interface. (**a**,**b**) STM images of **GC7** (1.1 x 10^{-6} M; I_{set} = 60 pA, V_{bias} = -300 mV). The unit cell is indicated by black lines (a = 2.5 ±0.1 nm, b = 4.6 ±0.1 nm, α = 106 ±1°). White lines indicate the normal axes of graphite. (**c**) Proposed association model with schematics of the stabilizing H-bonding motifs.

[225] a) P. Vanoppen, P. C. M. Grim, M. Rücker, S. De Feyter, G. Moessner, S. Valiyaveetil, K. Müllen, F. De Schryver, *J. Phys. Chem.* **1996**, *100*, 19636–19641; b) C.-J. Li, Q.-D. Zeng, C. Wang, L.-J. Wan, S.-L. Xu, C.-R. Wang, C.-L. Bai, *J. Phys. Chem. B* **2003**, *107*, 747–750; c) D. G. Yablon, D. Wintgens, G. W. Flynn, *J. Phys. Chem. B* **2002**, *106*, 5470–5475; d) K. Eichhorst-Gerner, A. Stabel, G. Moessner, D. Declerq, S. Valiyaveettil, V. Enkelmann, K. Müllen, J. P. Rabe, *Angew. Chem. Int. Ed.* **1996**, *35*, 1492–1495; e) P. Qian, H. Nanjo, T. Yokoyama, T. M. Suzuki, K. Akasaka, H. Orhui, *Chem. Commun.* **2000**, 2021–2022; f) H. Uji-i, M. Yoshidome, J. Hobley, K. Hatanaka, H. Fukumura, *Phys. Chem. Chem. Phys.* **2003**, *5*, 4231–4235.

By gradually increasing the monomer concentration in solution, more defects appeared in the domain (dashed black circles in Figure 68). These defects corresponded to the intercalation of cyclic tetramers between rows of already formed cyclic species, bound through the aminopyridine moieties. In more concentrated solutions, the network became more densely packed and the tetramers came closer together, gradually leading to a similar dense packing to the one observed for **GC3**. Unfortunately, more disorder was also observed in this case.

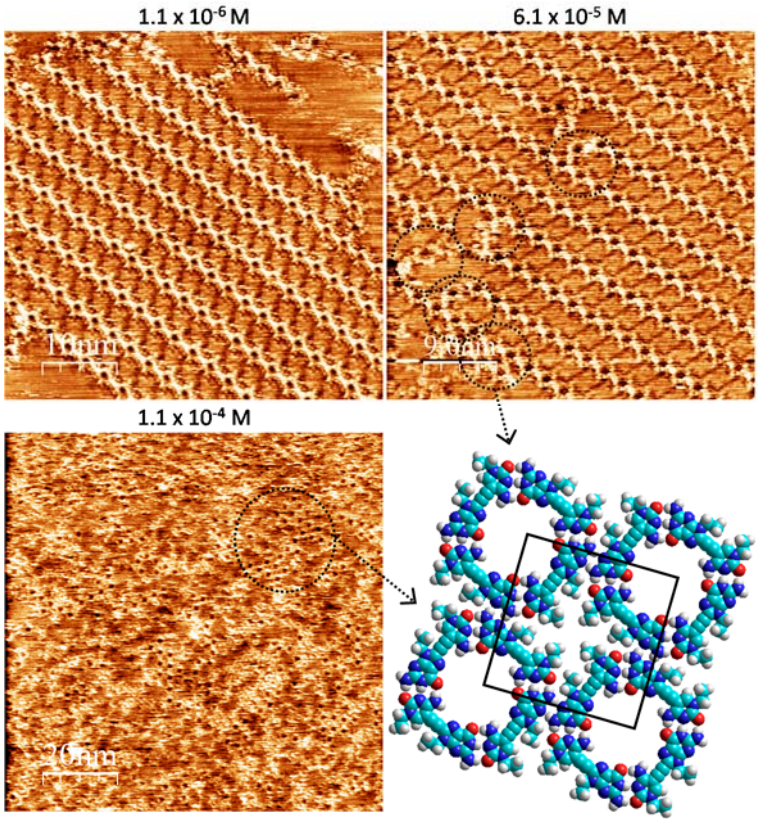


Figure 68. STM images of **GC7** for different concentrations at the HOPG/TCB:OA (1:1) interface (tunneling parameters: I_{set} = 60 pA, V_{bias} = -300 mV) along with model. The unit cell is indicated by black lines (a = 2.5 ±0.1 nm, b = 2.5 ±0.1 nm, α = 91 ±1°).

At this point, further one-component experiments without or with lower amounts of OA will be carried out in order to see if more densely-packed networks can be obtained. Annealing experiments might be another strategy that can be followed to explore the possibility of more densely-packed assemblies. Kinetics should also be investigated in order to analyze if a condensation of the network might be achieved with time. Additionally, host-guest studies with **GC7** and phenalene or C_{60} will be performed.

Medium-sized pores.

In section **2.2.2.**, a well-structured 2D nanoporous system formed by **GC3** presenting medium-sized pores, able to host molecules of an intermediate size, like **cor**, was described (Figure 60). However, since the one-molecule system suggested an arrangement where the pores were never totally empty, but filled with octyloxy alkyl chains, the possibility that the host-guest system was not at its most stable form arose, due to desorption of some alkyl tails.

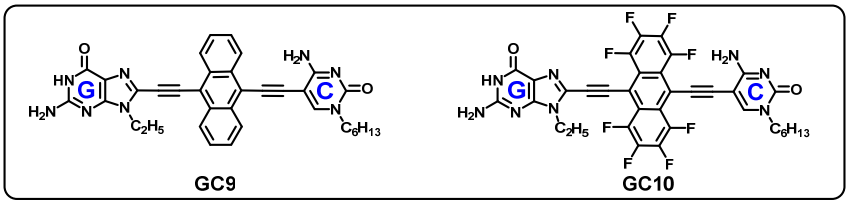
GC4. **GC4** one-component and host-guest systems will be soon studied by STM at the solid-liquid interface.

Large-sized pores.

GC8. The **GC8** one-component and host-guest systems with naphthalocyanine molecules (Figure 65) will be carried out in the near future by STM at the solid-liquid interface.

2.2.5. Functional Blocks: tuning monomer electronic properties.

A large family of central blocks can be incorporated between the self-assembling nucleobase directors in our supramolecular design (see the Background and Objectives section). In this manner, the monomers can be endowed with diverse functions by simply coupling the ethynyl nucleobases with a suitable molecule. In this section, we intend to explore the behavior of molecules with different functional blocks and see how the different functionalities affect their observation by STM. As these images can be considered a map of electronic density, two molecules with different overall electron properties have been considered to address this issue (Scheme 16).



Scheme 16. GC monomers for the study of functional systems.

Molecule **GC3** has shown its ability to self-assemble onto HOPG as large domains of cyclic tetramers stabilized through Watson-Crick selective H-bonding. Here, **GC3** has been taken as model, but instead of a phenyl ring equipped with long alkyl chains, a species with higher π-electron density has been incorporated. Candidate **GC9** is composed by an anthracene moiety substituted at one end with an ethyl-guanine derivative and at the other with a hexyl-cytosine. Molecule **GC9** has, therefore, a larger π surface and should exhibit a stronger affinity for the HOPG substrate than **GC3**. Besides, being more electron dense, **GC9**

should display a different brightness on the STM images. Target monomer **GC10** is a carbon copy of **GC9** but with electron-withdrawing fluorine atoms surrounding the central block. Molecule **GC10** should also appear with a different brightness on the STM images. In spite of the different electronic densities of the central block, the capacity of the monomer to self-assemble into tetrameric species should not be strongly affected.

GC9. A building block with high electron density is present in this molecule: the anthracene. The electron density difference between the different residues of the monomer translated into the molecule appearing now on the STM as a shiny dot, instead of a rod as the previous monomers **GC1** to **GC7**. By including anthracene as the rigid central block of the monomer, an enhanced affinity of the monomer for the HOPG was expected and therefore a very low molecular mobility on the surface was noticed, facilitating the submolecularly resolved characterization of the patterned surface by STM. Two different motifs have been observed: the so-called 'tetramers' and 'hexagons'. The first one was represented by four shiny dots forming a square (Figure 69b) and the latter by six combined dots forming a star-like structure (Figure 69d). In the different set of experimental conditions tested both patterns coexisted (Figure 69a).

Tentatively, each bright spot can be assigned to the location of an anthracene triple ring. Not all the dots appeared with the same brightness, indicating that not all the molecules had the same orientation with respect to the graphite substrate. Indeed, different tilt angles of the aromatic rings with respect to the substrate will result in differences in contrast.

The "tetramer" pattern corresponded to the desired cyclic tetrameric species formed by four molecules and bound by secondary H-bonds (Figure 69c). Based on previous observations on related GC molecules, secondary H-bonding interactions have been located between aminopyridine fragments in the models. However, since each monomer was seen as a circular shape, this cannot be seen in the images and other kind of secondary binding could also take place. Cyclic tetramers could also interact through their corners by establishing H-bonding interactions between the G-proton and the external C-carbonyl lone pair that do not participate in Watson-Crick pairing. In that case, a similar arrangement to the one afforded by **AU2** monomers (see below in Figure 72) could take place.

According to our models, the "hexagon" arrangement corresponded to the self-assembly of six cyclic trimers. Three **GC9** molecules would stablish distorted Watson-Crick interactions between their self-assembling units and form a cyclic trimer (Figure 69e). The resulting angle between nucleobases would be of 60° instead of the non-strained 90° angle exhibited by the cyclic tetramer. Despite not being preorganized for trimerization, **GC9** trimeric species is the smallest H-bonded macrocycle and is hence entropically favored. Secondary H-bonding interactions thorough the corners would lead to the observed porous honeycomb structure (Figure 69e). The appearance of this pattern could be explained by the lower mobility of the molecules that could "freeze", in the STM measuring time-scale,

this stable intermediate pattern in the formation of the more thermodynamic stable cyclic tetramer network.

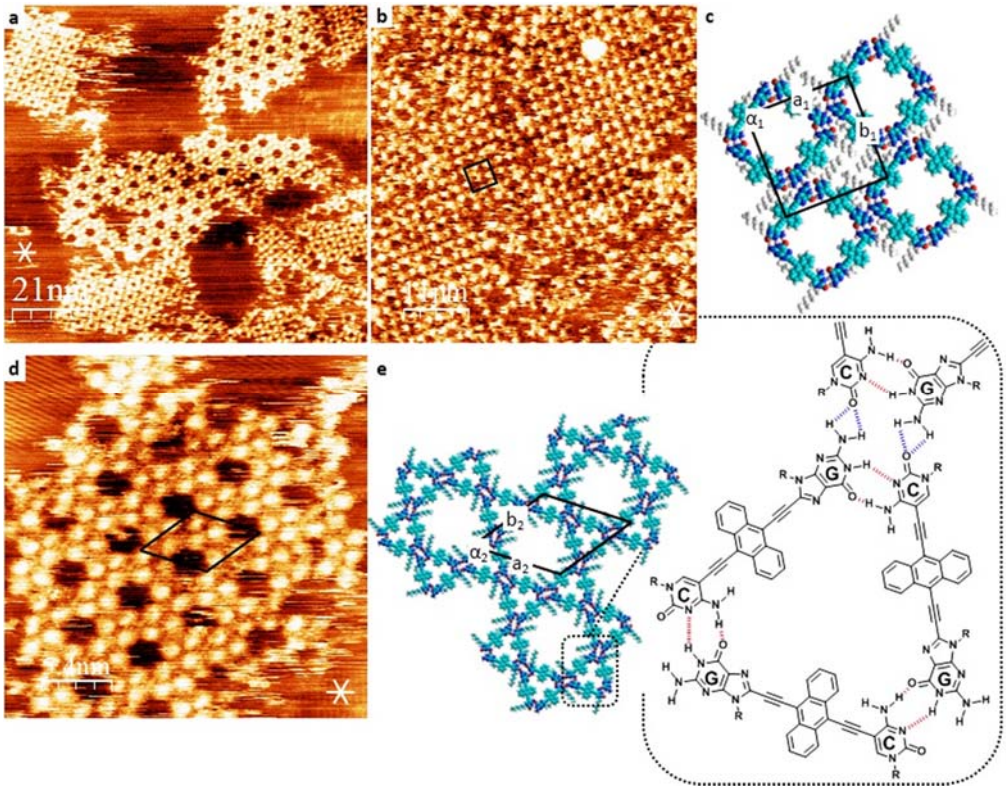


Figure 69. Self-assembly of GC9 on HOPG from a solution in TCB/OA (1:1). White lines indicate the normal axes of graphite. (**a**) STM image showing the coexistence of two patterns (1.80 x 10^{-4} M; I_{set} = 380 pA, V_{bias} = -150 mV). (**b**) Magnified STM image of a domain corresponding to the "tetramer" pattern (1.80 x 10^{-5} M; I_{set} = 300 pA, V_{bias} = -100 mV) and (**c**) its association mode. The unit cell is indicated by black lines (a_1 = 3.4 ±0.1 nm, b_1 = 3.3 ±0.1 nm, α_1 = 90 ±1°). (**d**) Magnified STM image of a domain corresponding to the "hexagon" pattern (1.20 x 10^{-5} M; I_{set} = 300 pA, V_{bias} = -150 mV). (**e**) Proposed association model with scheme of the stabilizing interactions. Black lines indicate the unit cell (a_2 = 4.6 ±0.1 nm, b_2 = 4.7 ±0.1 nm, α_2 = 60 ±1°).

Comparing the network dynamics of **GC3** and **GC9**, a higher nucleation phenomena with respect to the growing of the domain was also observed for **GC9**. In a large area this molecule presented numerous small nuclei (Figure 70). It is worth highlighting at this point the exceptionally strong molecule-substrate interactions displayed by monomer **GC9**. Typically, HOPG do not offer strong surface-molecule interactions and in order to stabilize an assembly onto this substrate, the molecules usually need to stablish numerous unions so that their cooperative action would result in an immobilization of the molecules in the STM time-scale. Small nuclei are thus not often observed by STM at the solid-liquid interface onto HOPG. By increasing the concentration, the coverage improved, and the

"tetramer"/"hexagon" pattern ratio increased, but also did disorder. Deposition of the solution into a hot substrate allowed the domains to grow. Let's remark that in these annealed STM images, one could almost exclusively contemplate the "tetramer" pattern. This was in concordance with this arrangement being the most thermodynamically stable and that the honeycomb pattern represents a kinetic intermediate in the self-assembly process.

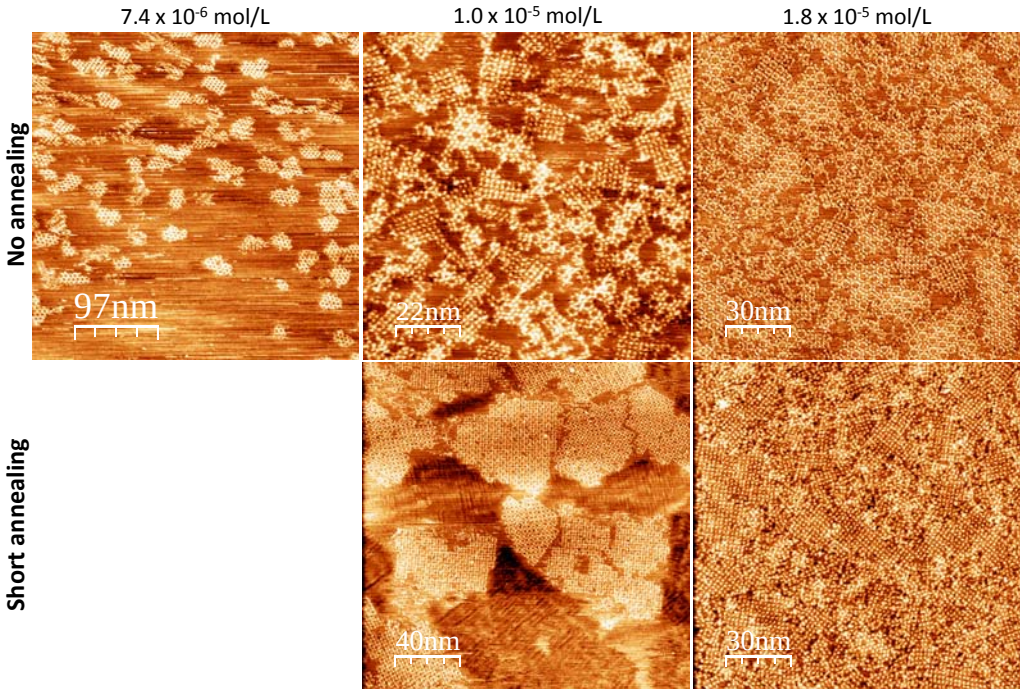


Figure 70. STM images of **GC9** at the HOPG/TCB:OA (1:1) interface (tunneling parameters: I_{set} = 100 pA, V_{bias} = -300 mV) at different concentrations and after no annealing or a short annealing process (80 °C; 30 min).

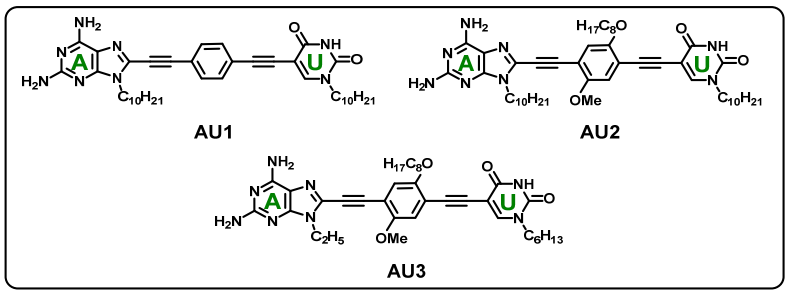
GC10. The preparation of monomer **GC10** was attempted by two consecutive Pd-catalyzed Sonogashira reactions between **C1$_{Alk6}$**, **G1$_{Alk2}$** and the synthesized central block **B5** (see Chapter 1). First, the optimized standard conditions for the Sonagashira-coupling reaction for these molecules were used to couple the dichlorooctafluoroanthracene block with the ethynyl-cytosine derivative. The dichlorinated central block was proven to be much less reactive than its dibromoanthracene analog in the Sonogashira reaction conditions. Still, the intermediate mono-coupled product **C1$_{GC10}$** could be isolated (see section **2.4.2.** in the Experimental Section). On the contrary, monomer **GC10** could not be obtained in these same conditions. In another attempt, $Pd(PPh_3)_2Cl_2$ was used as Pd-source, together with CuI, in the final coupling reaction with the G nucleobase in a DMF:diisopropylamine 1:1 mixture at 60 °C. This time, the starting products **G1$_{Alk2}$** and **C1$_{GC10}$** were consumed to afford a product, presumed to be **GC10**. Unfortunately, this product could not be properly purified

and characterized. Further optimization of the synthetic route will be carried out and the effect of the electron-withdrawing fluorine atoms will be studied by STM in next experiments.

2.2.6. A-U Cyclic Tetramer Networks: towards self-sorted systems.

The AU system.

In order to explore self-sorting phenomena on the surface with our system, networks exploiting other complementary nucleobases, like the A-U pair, need to be considered and their self-assembly into well-ordered cyclic tetramers through specific Watson-Crick has to be mastered first. In this section, we will study the AU cyclic tetramer system as the first step before designing self-sorting experiments. One more time, for the sake of simplicity, the focus was put on unsymmetric target molecules, assembling in one-component cyclic tetramers. Three different AU molecules (**AU1**-**AU3**), based on the structure of monomers **GC1** to **GC3**, have been synthesized (Scheme 17) with the aim of generating stable 2D nanoporous networks onto HOPG, formed by large domains of AU cyclic tetramers.



Scheme 17. Chemical structure of the target AU monomers.

AU1 is equivalent to **GC1** and carries 2,6-diaminopurine and uracil derivatives substituted with decyl chains, separated by a phenyl ring. **AU2** shares a common structure with **GC2** and the central block is unsymmetrically *p*-substituted with methoxy and octyloxy groups. **AU3** is a third generation monomer and is equipped with the same central block than **AU2** but the alkyl chains have been shortened in order to optimize the adsorption of the cyclic tetrameric species on the surface and maximize its stability, as seen for **GC3**.

AU1. Monomer **AU1** could not be stabilized on the surface and only some STM images corresponding to disordered molecules were recorded. Since the incorporation of substituents displaying affinity for HOPG (long alkyloxy chains, bigger π-surface) in the central block seemed crucial for obtaining GC macrocycles, a similar modification in the design of the monomer, like in the case of **GC1** molecules, wanted to be adopted. Therefore, we followed a similar strategy to the GC series and prepared monomer **AU2** with a central block substituted with a long octyloxy chain in order to partially fill the pores, so as to increase the surface density when the molecules self-assembled into cyclic tetramers.

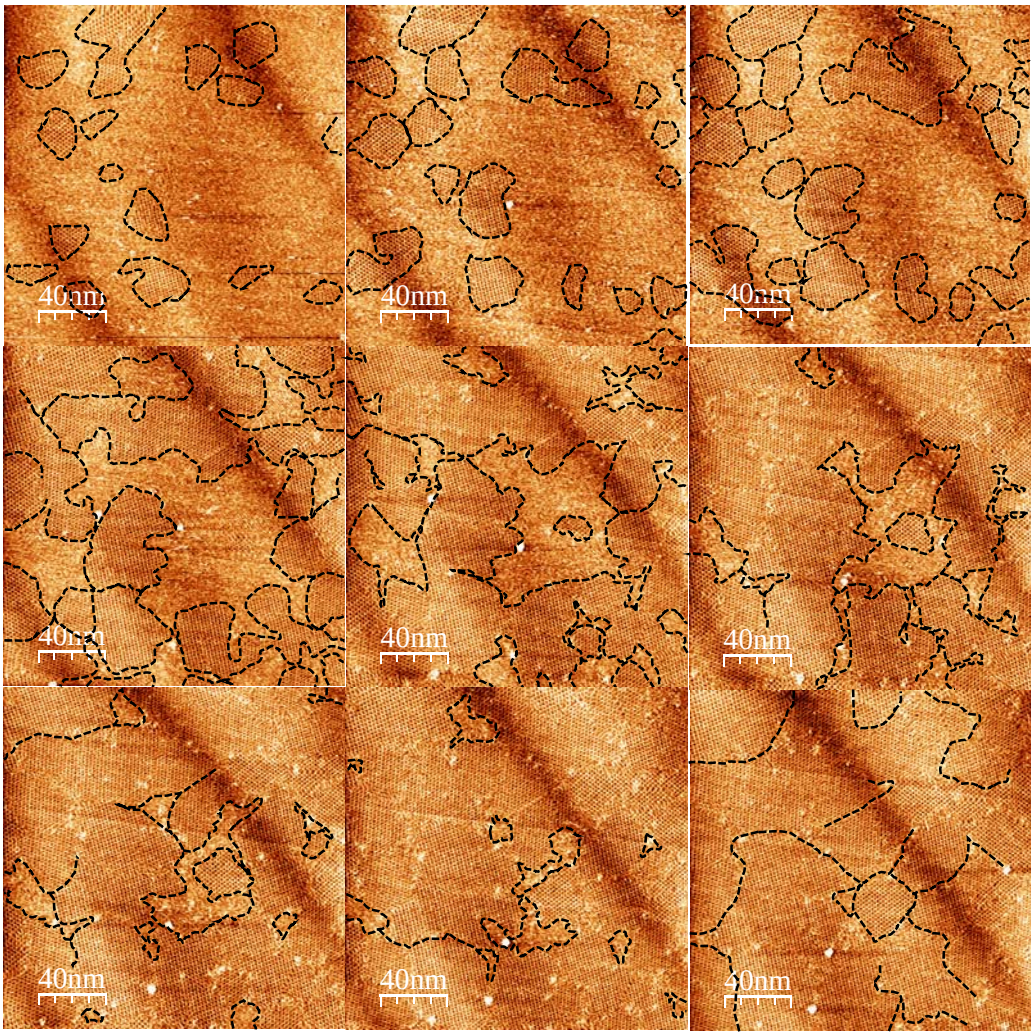


Figure 71. Successive scans over a 45 min period of a large domain (200 x 200 nm^2) after deposition of **AU2** in TCB:OA (1:1) on HOPG (7.9 x 10^{-5} M; I_{set} = 100 pA, V_{bias} = -350 mV). The evolution of the surface coverage can be appreciated. The black dashed lines mark domain borders.

AU2. The substitution of the central block with respect to **AU1** was efficient and well resolved images were obtained. A large area of medium-sized domains that grew with time until total coverage of the surface could be observed (Figure 71). Some noisy streaks could be appreciated in the borders of a condensed island, indicating that the molecules exhibited mobility on the time scale of one scan line. These domains evolved in time as shown by time-lapse imaging sequences. The borders of the 2D-condensed phase moved as a function of time, suggesting that the monolayer was in dynamic coexistence with the liquid phase and adsorbed disordered phases. These two phases coexisted in a 2D thermodynamic equilibrium at room temperature.

A fishnet-like nanoporous network formed by cyclic tetramers was obtained by drop-casting a 7.9 x 10^{-5} M solution of **AU2** in TCB:OA (1:1) onto HOPG (Figure 72). A closer look at the highly organized domains indicated that **AU2** displayed a distinct network arrangement than **GC2** or **GC3** (Figure 57). We found out that this is due to different secondary interactions between cyclic tetramers. Whereas **GC2** and **GC3** established G-G double H-bonding interactions between aminopyridine-type fragments, leading to regular lattices in which the rings interacted through their G-edges; the squared-shaped **AU2** cycles bound through their corners by establishing H-bonding interactions between the A-H^2 proton and the external U-carbonyl lone pair, so the resulting network appeared as a continuous grid.

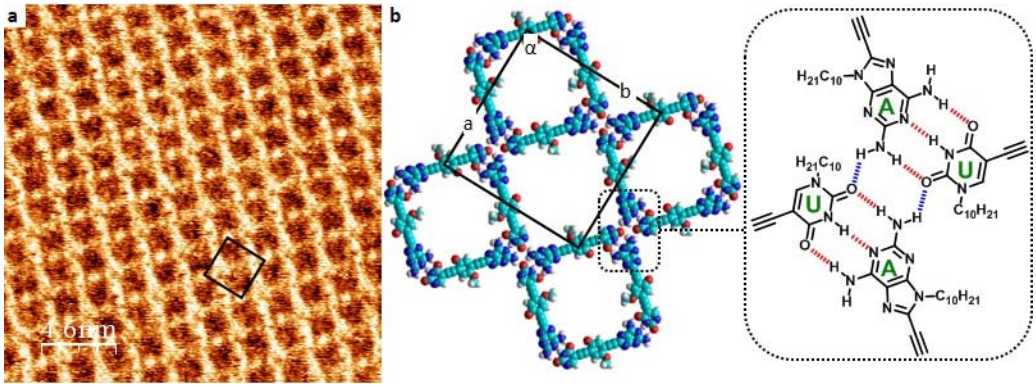


Figure 72. AU2 self-assembly. (**a**) STM image of **AU2** on HOPG from a solution 7.9 x 10^{-5} M in TCB:OA (1:1) (tunneling parameters: I_{set} = 100 pA, V_{bias} = -350 mV). The unit cell is indicated by black lines (a = 3.5 ±0.1 nm, b = 3.5 ±0.1 nm, α = 91 ±1°). (**b**) Proposed association model with scheme of the stabilizing H-bonded motif. Dashed red lines indicate Watson-Crick base-pairing and the dashed blue lines correspond to secondary H-bonding between nucleobases.

It should be remarked that **GC2**/**GC3** or **AU2** could in principle establish any of these two secondary H-bonding configurations (Figure 73), but these distinct centrosymmetric lattice arrangements were always observed for each molecule under different concentration conditions and annealing or time-dependent studies did not show any network reorganization, suggesting that they constitute equilibrium states. Both network arrangements onto HOPG left secondary pores between cyclic tetramers, having a comparable size than the ones defined by the macrocycle cavity, where the external alkyl tails placed at the nucleobases were forced to group. The interaction between base pair corners left secondary pores that were significantly larger (5.6 nm^2) than those generated by interaction between purine edges (3.4 nm^2), where the long peripheral tails should have fitted better and established stronger interactions with the substrate. As a matter of fact, larger and better-ordered domains were observed for **AU2**, for which the surface density was calculated as 232.1 g/mol·nm^2, than for **GC2** (Figure 56). Still, the reasons why **GC2** tetramers prefer to bind through the purine edges are not clear. Our hypothesis, developed in Figure 73, is that this particular arrangement may be a consequence of short-range

cancelation of the nucleobase dipole moments, which are much stronger in the G-C pair than in the A-U pair.[226] Nevertheless, our results reflected the subtle interplay between supramolecular forces when molecules are packed in two dimensions.

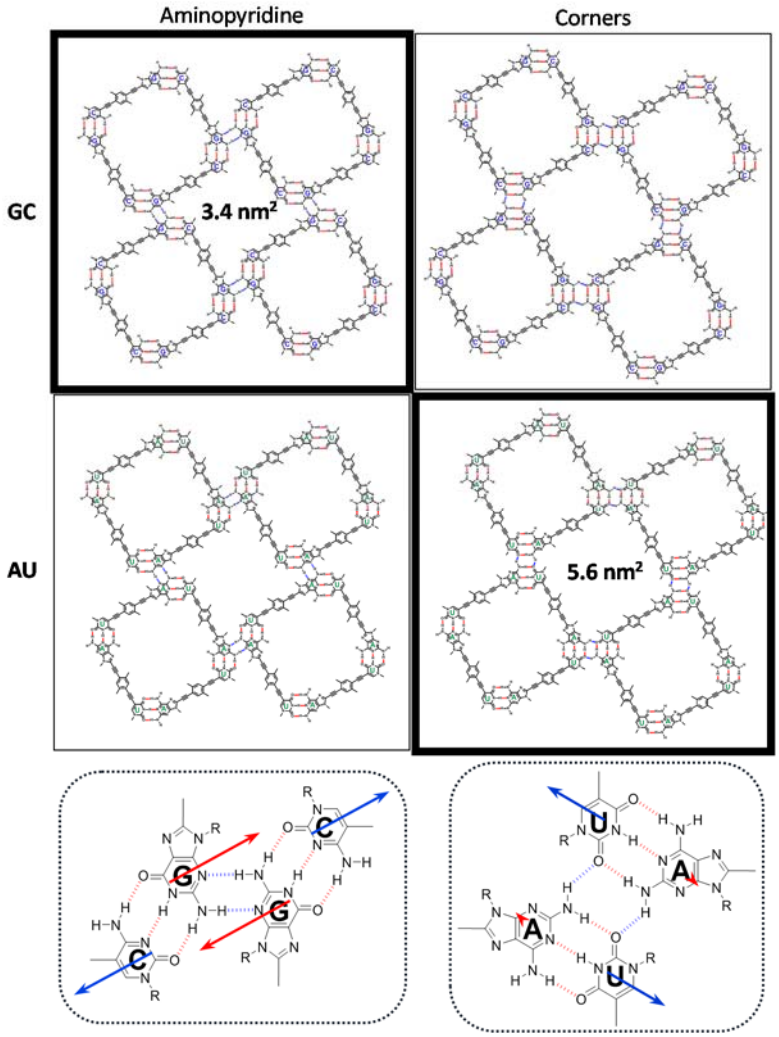


Figure 73. GC and **AU** cyclic tetramers secondarily H-bonded through their purine aminopyridine edges or through their corners. The darker squares indicate the kind of arrangement that is always observed experimentally for each molecule. The secondary pores between cyclic tetramers are significantly smaller when the network grows by interaction between the purine edges (3.4 nm^2) than when the pairs of bases interact through their corners (5.6 nm^2). The arrows indicate the magnitude and direction of the molecular dipole moments of each nucleobase,[226] which are much stronger in the G-C pair than in the A-U pair. All H-bonding patterns are centrosymmetric but short-range cancelation of the strong G dipoles may be more efficiently achieved *via* aminopyridine interactions.

[226] J. Sponer, J. Leszczynski, P. Hobza, *Biopolymers*, **2001**, *61*, 3–31.

In view of these results, we decided to prepare a third AU monomer (**AU3**) with shorter $–C_2$ and $–C_6$ chains, whose synthesis was carried out from the corresponding A and U derivatives. The same reasoning followed for GC monomers has been extrapolated to the AU family. What we want to see is whether monomer **AU3** keeps stablishing secondary H-bonding interactions through the corners, like **AU2**; or whether the network can be rearranged through the interaction between A-aminopyridine fragments, like the GC series, since the shorter chains $–C_2$ and $–C_6$ fit perfectly in the space left between cyclic tetramers in this network. Monomer **AU3** will afford valuable information for the better understanding of the self-assembly processes of these molecules on the surface.

AU3. **AU3** has not been studied yet by STM at the solid-liquid interface, but will be considered in next experiments.

Self-sorting experiments.

As introduced earlier in the Background and Objectives section, by self-sorting we intend a network where pure AU and GC cyclic tetramers can be segregated on the surface and imaged separately by STM (Figure 74b). The interactions between nucleobases through their Watson-Crick edges are specific and self-sorting in AU_4 and GC_4 tetramers is likely to occur, as demonstrated by our group in solution experiments (see Background section). However, we should distinguish an additional level of self-sorting on the surface that may occur at a larger nanodomain scale (Figure 74d,e). The second-order hierarchical H-bonding that builds the network, either *via* the macrocycle's edges or corners, may take place between AU_4 tetramers, GC_4 tetramers or between AU_4 and GC_4 tetramers, since they both have identical H-bonding donor and acceptor groups at the tetramer periphery (see Figure 73).

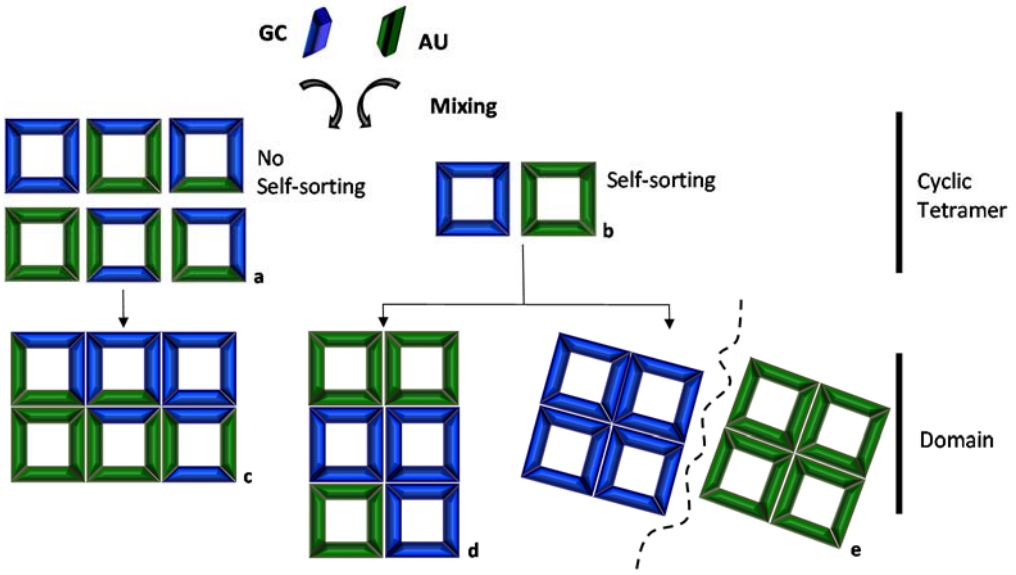
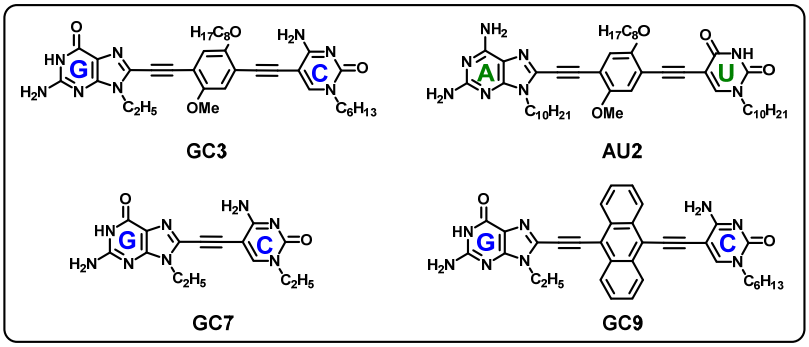


Figure 74. Schematic representation of self-sorting phenomena at the cyclic tetramer and domain levels. (**a**) The system do not self-sort and mixed cyclic tetramers are obtained. (**b**) The system self-sorts and pure GC and AU macrocycles are obtained. (**c**) A network of mixed tetramers formed by GC and AU monomers is observed. (**d**) There is no phase segregation and pure GC and AU cyclic tetramers form a single domain. (**e**) A network with separate domains of GC tetramers, on the one hand, and AU tetramers, on the other, is obtained.

The problem to differentiate a GC monomer from a AU monomer on the STM images would arise at this point. Notwithstanding, the versatility of the approach in our hands grants a variety of resources to prepare molecular targets with characteristic features on the STM images. First of all, the resolution of the STM apparatus allows, unquestionably, to distinguish between molecules of different sizes, with an error of 0.1 nm. Consequently, one easy way to differentiate molecules carrying different Watson-Crick nucleobase families would be to dropcast on the substrate a short GC monomer, like **GC6** or **GC7**, and a longer AU monomer, like **AU2** or **AU3**. Let's specify that mixed secondary binding between AU_4 and GC_4 tetramers should be unlikely in this case, as the created net would result uneven and the network packing could not be optimal. Another possibility for the visual differentiation of monomers is to consider that STM images are a map of electronic density. We have already seen in the previous section **2.2.5.** that monomer **GC9** has a different appearance in the images on HOPG, in comparison to the other molecules equipped with ethynyl or phenyl derivatives as central blocks. Therefore, **GC9** could be potentially distinguished by STM, from all the other GC or AU monomers in the mixture. In order to study these mixed systems and taking into account the aspects exposed above, the following family of monomers was proposed for self-sorting studies on 2 dimensions (Scheme 18).



Scheme 18. Considered AU and GC monomers for the study of mixed systems and self-sorting.

Different GC monomers (*i.e.* different length, different electron density) have been considered, allowing the possibility to tell apart on the STM images one GC monomer from another, or from a AU monomer. The single components may have a different adsorption energy and each of them have a tendency to minimize the occupied areas. Therefore, control of the deposition of two different molecules on a surface can be considered a rather difficult task. In order to check nucleobase complementarity and the Watson-Crick base-pairing model in our molecular system, the following experiments have been designed (Figure 75).

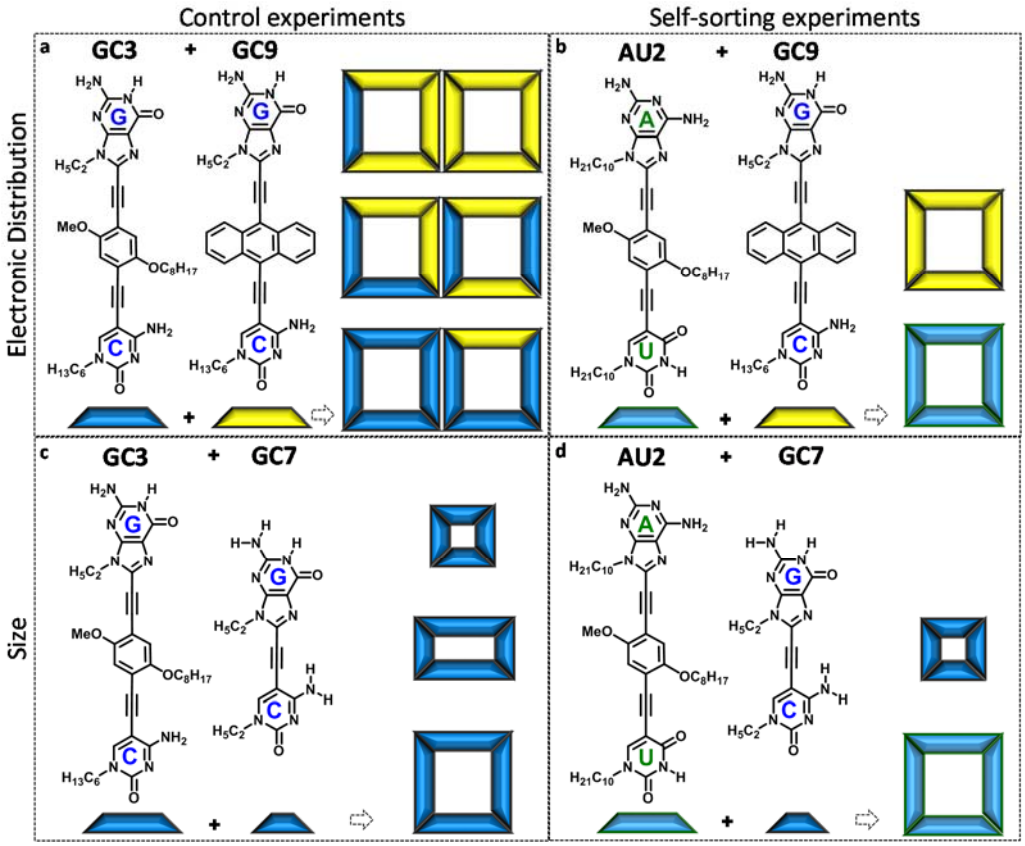


Figure 75. Proposed control and self-sorting experiments based on (**a**,**b**) electronic density, (**c**, **d**) size differentiation of our molecular system for the elucidation of self-sorting phenomena on the surface. **GC9** different appearance on the STM images is schematically represented by the yellow color.

As we have seen before, two monomers can be differentiated in the STM images either by their size, shape or electronic distribution; and this can be useful to design valid strategies to prove our hypotheses. In the self-sorting experiments shown In Figures 75b and 75d, we expect to be able to differentiate pure AU from GC tetramers. On the one hand, two families of nucleobases are present in the mixture (**b**) and thus, according to base-pairing complementarity, distinct GC and AU cyclic tetramers are expected. Since secondary H-bonding interactions are unspecific, the STM images could show both mixed and segregated domains. Monomer **GC9** should appear as a characteristic bright sphere (see section **2.2.5.**) and **AU2** as a rod with moderate brightness. On the other hand, by co-depositing **GC7** and **AU2** on HOPG, one could check if the system self-sorts by identifying domains of small nanopores and domains of bigger cavities (Figure 75d). However, in both cases (**b** and **d**), one would have to differentiate between phase segregation phenomena and self-sorting due to nucleobase complementarity. Therefore, in order to validate the possible positive results from experiments (**b**) and (**d**), one control experiment for each

approach (Figure 75a,c) has been designed as well. In the control experiment (**a**), both **GC3** and **GC9** molecules have the same length but differ in the electronic distribution. As seen in section **2.2.5.**, the two molecules have different appearances on the STM images on their own and normally the two components should be distinguishable on the images of the mixture. Since both of them are GC monomers, each nucleobase could bond either with the complementary nucleobase from the same kind of molecule or from the other one. The second control experiment (**c**) would consist on drop-casting a mixture of monomers **GC7** and **GC3** to see if three tetramers with different size and shape are obtained. The only possible three species are **GC3**$_4$, **GC7**$_4$ and **GC3**$_2$**GC7**$_2$, where two **GC3** monomers are placed between two opposed **GC7** molecules and a rectangular-shaped cyclic tetramer would be obtained. Monomer **GC7** is shorter than **GC3** and this time, a random distribution of cyclic tetramers within the same domain is not expected as the secondary H-bonded packing would be uneven, resulting in an unstable network.

GC3+GC9 *vs* AU2+GC9. The control mixture **GC3**+**GC9** was checked first (Figure 76a). In theory, the two monomers should not show a strong preference in their binding and a statistical mixture of six different tetramers was expected. Also, each tetramer could be secondarily bound to any other tetramer and consequently, the **GC3**+**GC9** combination would, in principle, afford a very complex mixture. Nevertheless, the stability of each molecule (or each cyclic tetramer) onto HOPG should not be the same (as seen in sections **2.2.1** and **2.2.5.**) because of different adsorption energies, as a function of the size of the aromatic central unit. This means that even if a statistical mixture is in theory expected, the STM images might not show evidence of this, despite the lack of a preferential binding between nucleobases.

After testing different mixtures, good imaging was achieved with a **GC3**:**GC9** 2:1 diluted solution, in the conditions indicated in Figure 76. STM images showed total coverage of the scanned area by tetrameric cyclic species, proving the stability of the system. However, the domains were not as large as for monomer **GC3** alone. Moreover, the shape and appearance of the cyclic entities was not the same as that encountered for **GC3** or **GC9** separately, suggesting that the monomers did mix. It is interesting to note that no “hexagon” pattern was observed in these mixtures (see Figure 69). In these images, tetramers formed by two different monomers could be observed. Unfortunately, the differentiation of each monomer in the tetramer was not as straightforward as expected. A tentative assignment of this complex mixture is showed in Figure 76c.

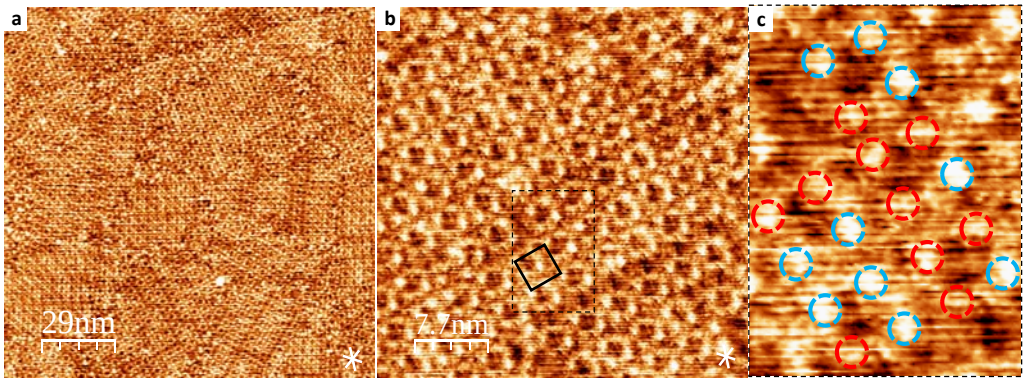


Figure 76. STM images of a premixed solution (2:1) of GC3 (6.0 x 10^{-6} M) and GC9 (3.0 x 10^{-6} M) in TCB:OA (1:1) on HOPG (tunneling parameters: I_{set} = 60 pA, V_{bias} = -420 mV). White lines indicate the normal axes of graphite. (**a**) Large scale STM image showing the total coverage of the surface by numerous medium-sized domains. (**b**) Small scale STM image indicating the unit cell by black lines (a = 3.3 ±0.1 nm, b = 3.3 ±0.1 nm, α = 91 ±1°). (**c**) Magnified STM image corresponding to the delimited area in dashed black lines in (**b**) and tentative scheme of mixed tetramers (color red indicates **GC3** and blue **GC9**).

This control experiment was considered as successful, since no self-sorting, nor phase segregation, seemed to occur. The obtained results were encouraging in the sense that two important points had been proven, granting validity to the next experiments. First of all, one can stablish that molecular recognition between nucleobases drives the self-assembly process into cyclic tetramers. G nucleobases can bind to any C-nucleobase and, therefore, in the mixture of two GC-monomers no self-sorting occurs. Secondly, despite the difference in stability of the two molecules onto the substrate, apparently no phase segregation took place and the dynamics of the mixture was intermediate between the pure **GC3** and **GC9** networks. This control experiment denoted importance because it showed that phase segregation phenomena are not dominant in the stabilization of the monolayer on the surface. Therefore, if different domains of AU and GC cyclic tetramers were to be obtained after drop-casting a AU+GC mixture of monomers onto HOPG in a second experiment, one could come to the conclusion that the system self-sorts. The outcome would be a result of H-bonding specificity and base-pairing complementarity and not an artifact of phase segregation.

Having obtained favorable results on this control experiment, the **AU2**+**GC9** mixture was considered at this point. Different premixed solutions were prepared taking into account the defined optimal concentrations for imaging the cyclic tetramer networks observed when studying the one-molecule systems. Solutions containing 1:1 to 1:10 ratio **GC9**:**AU2** mixtures were essayed. Unfortunately no good imaging of this **GC9**+**AU2** system was achieved. It is possible that a short annealing process could help reaching equilibrium conditions for the imaging of a stable network. Further optimization of the experimental

conditions will be carried out in the future in order to stabilize a well ordered arrangement of, hopefully, appreciable self-sorted cyclic tetramers onto HOPG.

GC3+GC7 *vs* AU2+GC7. We focused then on the second strategy, this time based on size differentiation. In the **AU2**+**GC7** mixture (**d**), only two kind of cyclic tetramers were predicted: small squared GC-tetramers and medium-sized squared AU-tetramers. No rectangular-shaped tetramers were expected. Since **GC7** monomer is shorter than the **AU2** monomer, both should be easily recognizable on the STM images. However, in this mixed system, the different adsorption of each of the molecules onto HOPG resulted problematic. The concentration required to image correctly each monomer on its own differ vastly from one molecule to the other. When fixing the concentration of each of the monomers according to the optimized conditions seen for the one-component systems, **AU2** resulted in large excess (40:1). Consequently, in these conditions the characteristic medium-sized AU cyclic tetramer grid-like network was exclusively imaged by STM (Figure 77a). When increasing the concentration in **GC7**, too much disorder was observed in the STM images (Figure 77b). Despite the fact that no large ordered domains of cyclic tetramers were obtained in these disordered images, small aggregates of recognizable medium-sized AU and small-sized GC cyclic tetramers could be distinguished (Figure 77b).

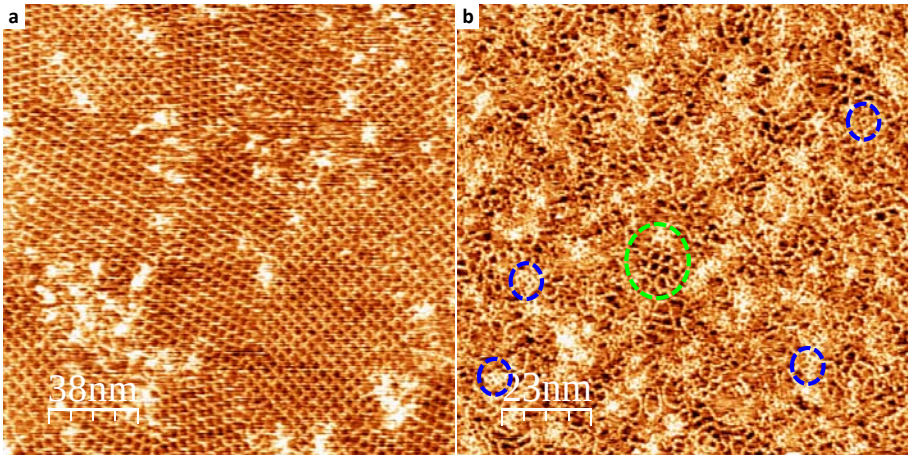


Figure 77. STM images of premixed solutions in TCB/OA (1:1) of AU2 and GC7 on HOPG (tunneling parameters: I_{set} = 80 pA, V_{bias} = -300 mV). (**a**) Premix in a 40:1 ratio (**AU2**: 4.0 x 10^{-5} M and **GC7**: 1.1 x 10^{-6} M). (**b**) Premix in a 1:3 ratio (**AU2**: 4.0 x 10^{-5} M and **GC7**: 1.2 x 10^{-4} M). Blue dashed circles indicate small GC tetramers and green AU bigger macrocycles.

These self-sorting experiments require further optimization and additional experiments will be performed. Annealing processes or simply letting the mixture stand during longer periods so as to allow the sample to reach the thermodynamic equilibrium might help in obtaining larger domains.

The mixture of monomers **GC3** and **GC9** (**a**) showed that no phase segregation was taking place as combined cyclic tetramers were obtained (Figure 76b). However, the control

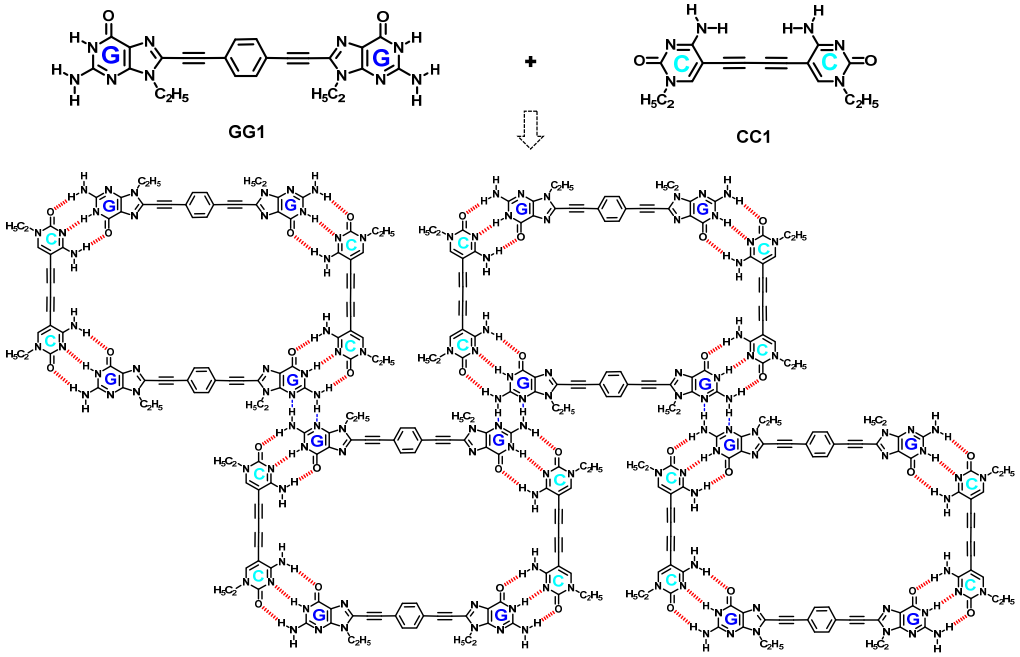
experiment (**c**) should be carried out as well, in order to confirm that phase segregation phenomena are not dominating in the self-sorting experiment (**d**). Unfortunately, the **GC3**+**GC7** mixture has not been studied yet and these measurements are still pending. This system will be considered in future STM experiments and will help gaining more insight into the behavior of this molecular system in relation to self-sorting and phase segregation phenomena.

As a next step, we would like to carry out dual recognition experiments with this system. Once the conditions are optimized so that a mixed solution of the third generation **AU3** monomer and **GC7** self-sorts to give two different domains with distinct pore sizes, easy to recognize by STM, the self-sorting experiments could be enlarged by studying the respective host-guest systems with appropriate size-complementary guests (the small pore would host phenalene and the medium pore **cor**, for instance). In these further experiments, GC and AU macrocyles could also be differentiated by guest inclusion. Actually, the construction of self-assembled crystalline networks that consist of three or more components are rare both under UHV conditions[95] and at liquid solid interfaces[102,227] and therefore, this project holds great interest.

2.2.7. Two-component Systems: toward pore shape control.

A final specific objective that wanted to be fulfilled was the discrimination of guest molecules as a function of their shape. Profiting again from our versatile supramolecular system, two different symmetric monomers (A and B) can be mixed in order to obtain ABAB cyclic tetramers with rectangular shape. Some ABAB systems have already been studied in solution by us (see the Background section). These two-component systems are now addressed on the surface. Two monomers with different length (**GG1** and **CC1**) have been designed (Scheme 19) to form a rectangular-shaped two-component cyclic tetramer. Nucleobases G and C have been chosen to maximize H-bonding strength and directionality in the Watson-Crick unions and further stabilize the cyclic species. The expected network is also shown in Scheme 19.

[227] a) J. Adisoejoso, K. Tahara, S. Okuhata, S. Lei, Y. Tobe, S. De Feyter, *Angew. Chem. Int. Ed.* **2009**, *48*, 7353–7357; b) J. Liu, T. Chen, X. Deng, D. Wang, J. Pei, L.-J. Wan, *J. Am. Chem. Soc.* **2011**, *133*, 21010–21015; c) K. Cui, F. Schlütter, O. Ivasenko, M. Kivala, M. G. Schwab, S.-L. Lee, S. F. L. Mertens, K. Tahara, Y. Tobe, K. Müllen, K. S. Mali, S. De Feyter, *Chem. Eur. J.* **2015**, *21*, 1652–1659.



Scheme 19. Targets **GG1** and **CC1** for shape discrimination studies and the expected cyclic tetrameric ABAB system.

The preparation of monomers **GG1** and **CC1** is detailed in the Experimental Part (**2.4.2.**) of this Chapter. Monomer **GG1** was prepared according to the general synthetic route explained in Figure 53, by coupling 2 equivalents of the $\mathbf{G1_{Alk2}}$ derivative to the dioodobenzene central block. Contrarily to all the other synthesized final products, which were obtained after two Sonogashira coupling reactions with $Pd(PPh_3)_4$ as palladium-source, homocoupling products are more efficiently obtained by using other reaction conditions.[228] However, target monomer **CC1** was obtained as a byproduct of the standard Sonogashira reaction conditions and it was purified afterwards.

In the proposed network (Scheme 19), the cyclic species are stabilized by H-bonding through the Watson-Crick faces of the nucleobases. The network is further stabilized by secondary interactions between aminopyridine fragments, allowing the formation of a densely-packed network. Alkyl chain lengths have been carefully designed taking into account this simulation: both nucleobases are equipped with $-C_2$ chains. Let's point out that, once again, unspecific cavities have been blocked.

First of all, the self-assembly of these molecular building blocks will be studied. If successful, host-guest shape discrimination experiments can be carried out where this

[228] a) D. Schindler, F. Eiβmann, E. Weber, *Org. Biomol. Chem.* **2009**, *7*, 3549–3560; b) Y. He, C. Cai, *Catal. Sci. Technol.* **2012**, *2*, 1126–1129; c) K. Yin, C.-J. Li, J. Li, X.-S. Jia, *Appl. Organometal. Chem.* **2011**, *25*, 16–20.

rectangular-shaped tetramer can host an elongated guest like pentacene but not a circular guest like **cor**, for instance (Figure 78).

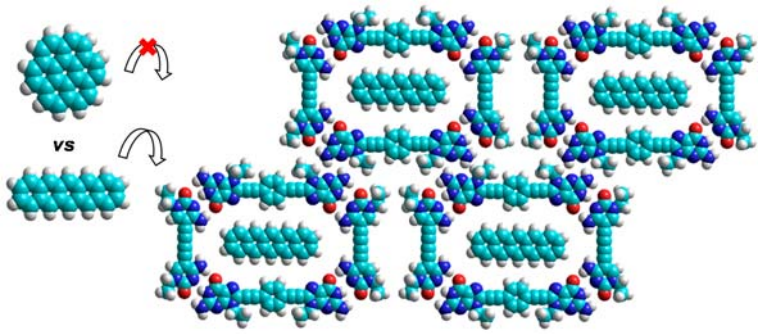
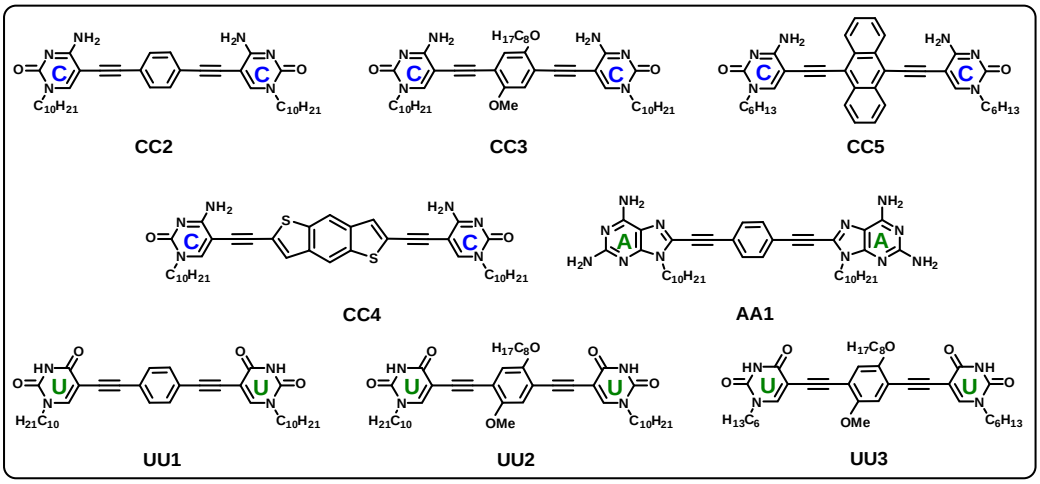


Figure 78. Proposed shape-discrimination experiment between ABAB tetramers (composed of **GG1** and **CC1**), and pentacene as shape complementary guest.

The following symmetric monomers (Scheme 20) have been isolated from previous reactions targeting the unsymmetric monomers studied in prior sections. Since they represent a useful collection for future studies of two-component ABAB systems, these products have been properly purified and characterized. This data can be found in the Experimental Section of this Chapter.



Scheme 20. Other symmetric CC, AA and UU monomers available for the investigation of bicomponent cyclotetramerization on the surface.

2.3. Summary and Conclusions.

A novel strategy based on molecular self-assembly toward unconventional 2D nanoporous systems has been described. This approach differs from previous H-bonded lattices where the porous network is built on the surface from weak interactions between di- or tritopic molecules (the phthalic or trimesic acids being representative examples). Here, a robust macrocycle was assembled from strong Watson-Crick interactions between DNA bases whose supramolecular structure can be preserved upon surface transfer by rational molecular tailoring.

A thorough strategy has been followed in order to stabilize such cyclic tetramers on the surface. First of all, the monomers were designed, then synthesized, and finally studied by STM at solid-liquid interface. Rigorous molecular tailoring has been carried out taking into account experimental results and applying them into new designs.The obtained results were carefully analyzed in order to correct and redesign the molecular structure of the monomeric candidates. Several generations of monomers have been prepared (**GC1** to **GC4** and **AU1** to **AU3**), each improving the properties of the previous one in order to achieve the stabilization of porous networks of cyclic GC and AU tetramers.

It was seen that the 2D cyclic tetramer networks were actually stabilized by hierarchical H-bonding interactions. Aside from the Watson-Crick pairs, distinct secondary interactions were identified in the stabilization of GC and AU tetramers from groups that were not participating in the Watson-Crick pairing. In the GC structures, H-bonding between aminopyridine fragments occurred, so the tetramers were bound through their G edges to form the network. In contrast, in the AU system, secondary H-bonding interactions were established between the external U carbonyl lone pair and A amine proton, so the AU tetramers were bound through their corners. A tentative explanation for this phenomenon has been proposed, where the strongest GC nucleobase dipole moment is more efficiently cancelled in the short-range “edges arrangement” than in the “corners organization” displayed by **AU2**. Further experiments are however needed in order to fully understand this behavior. For this purpose, molecule **AU3** was synthesized and will be studied by STM at the solid-liquid interface in order to gain more insight into the behavior of these hierarchical H-bonding interactions.

Apart from the stabilization of macrocycles on the surface, an ambitious and detailed investigation of molecular self-assembly at the solid-liquid interface has also been drawn up to explore other matters like size-discriminating host-guest systems, chirality on the surface, functionality, self-sorting, or two-component systems. A large collection of DNA-based ditopic molecules (22) was thus designed, synthesized, isolated and characterized in order to address each of the previously mentioned topics. Twelve unsymmetric monomers carrying different complementary nucleobases, with different size, functional block and endowed with a chiral center or not, were prepared. Ten of them were based on the

Watson-Crick G-C pair and three on the A-U pair. Also ten symmetric monomers substituted with the four main DNA bases were isolated.

The ability of the nanoporous network to host size-complementary guests was preliminary demonstrated and stable bimolecular assemblies with coronene were reproducibly obtained. Dual recognition (first between the G and C nucleobases; and then, between the cyclic host assembly and the guest molecule) was demonstrated at this point.

Chiral domains have been assembled by the establishment of multiple and distinct H-bonding interactions, of a second-order hierarchy, between tetrameric macrocycles. A chiral monomer **GC5** was prepared to explore the possibility of reaching single pore chirality and an enantiomorphous nanoporous network with it, representing a good starting point toward surfaces potentially able to selectively host enantiomeric guest molecules.

Stable 2D networks of cyclic tetramers of different size and displaying different functionality were self-assembled onto HOPG. The insolubility of the monomers turned out to be an added difficulty to the self-assembly study by STM at the solid-liquid interface where not all the solvents are suitable. The mixture of TCB and OA solvents was proven to be adequate for the solubilization of the molecules. However, further optimization of the experimental conditions will be required in order to obtain more densely-packed networks, as solvent molecules may sometimes intervene in the self-assembly process on the surface. Dynamics seemed to play an important role in the self-assembly of these structures as well. Temperature and time parameters will be considered more carefully in the future in order to explore the behavior of these systems and see how they adapt when tuning different experimental conditions.

The first steps toward self-sorted and multicomponent systems have been given; as well as regarding size- and shape-discriminating host-guest systems. Unfortunately, not all the experiments by STM at the solid-liquid interface have been carried out yet. However, a wide set of molecules has been carefully prepared and will be further studied in the near future to fulfill these studies and pave the way toward molecular recognition at different levels.

2.4. Experimental Section.

The General Methods detailed in the Experimental Section **1.4.1.** of Chapter 1 are also applicable here. Some of the work described in this Chapter can also be found in the Supporting Information of our paper: *Angew. Chem. Int. Ed.* **2016**, *55*, 659–663 (ref.229).

2.4.1. Details on the STM measurements.

All the results reported in this Chapter were obtained by means of STM at the solid-liquid interface. All STM experiments were performed at room temperature (20–23 ^{o}C) using a PicoLE (Agilent) or PicoSPM (Molecular Imaging, now Agilent) machine operating in constant–current mode with the tip immersed in the supernatant liquid. STM tips were prepared by mechanical cutting from Pt/Ir wire (80%/20%, diameter 0.2 mm). Stock solutions of the monomers were prepared by dissolving approximately 0.1 mg of solid *per* milliliter of the pure solvent (1-octanoic acid or 1,2,4-trichlorobenzene) and in some cases in a 1:1 mixture of the two. The stock solutions (concentration around 10^{-4} M in most cases) were diluted to make concentration series. Commercially available coronene (Aldrich Co.), 1-octanoic acid (Sigma Co., 99%), and 1,2,4-trichlorobenzene (Sigma-Aldrich Co., 99%) were used without further purification. Such low vapor pressure solvents, (electro)chemically inert and with less affinity to form ordered monolayers on the surface than the molecules of interest, are appropriate liquids for STM imaging at the liquid-solid interface and are typically used. Other suitable solvents, like 1-phenyloctane, were not used because of compound insolubility reasons. The monomers were dissolved in the indicated solvent mixtures and the solutions were clear after 1 hour sonication. The concentrations used in each experiment are indicated in each case. Prior to imaging, a drop of the solution was applied onto a freshly cleaved surface of HOPG (grade ZYB, Advanced Ceramics Inc., Cleveland, USA). Short annealing processes were sometimes performed by heating the HOPG substrate to 80 ^{o}C and applying a drop of hot solution. STM imaging started immediately after drop casting. The experiments were repeated in several sessions using different tips to check for reproducibility and to avoid experimental artifacts, if any. For analysis purposes, recording of a monolayer image was followed by consecutive imaging the graphite substrate underneath. This was done under the same experimental conditions but by lowering the substrate bias (typically V_{bias} = 1 mV) and increasing the tunneling current (typical I_{set} = 800 pA). From the atomically resolved STM image of HOPG one can easily obtain the graphite symmetry axes. The images were corrected for drift *via* Scanning Probe Image Processor (SPIP) software (Image Metrology ApS), using the graphite lattice, allowing a more accurate unit cell determination. The unit cell parameters were calculated after analysis of 8 to 15 well-resolved and corrected images (30 x 30 nm^2). From the corrected graphite images three main symmetry axes and also three reference axes (normal

[229] http://dx.doi.org/10.1002/anie.201509233.

to main symmetry axes) can be determined. The images are low-pass filtered. The imaging parameters are indicated in each figure caption: tunneling current (I_{set}), and sample bias (V_{bias}). The molecular models were built using Hyperchem™ 8.0 program.

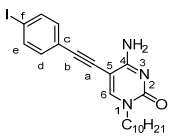
2.4.2. Synthesis and Characterization.

The synthesis and characterization of **B2** and **B5** central blocks, **C2$_{Alk2}$** and **C2$_{Alk10}$** precursors and the 5-/8- ethynylated lipophilic nucleobases were described in the previous Chapter 1 (section **1.4.2.**). The *Standard Procedure B* for the Sonogashira coupling with TMSA and subsequent alkyne-TMS group deprotection was detailed in section **1.4.2.** as well.

- ***Standard Procedure F*** for the Sonogashira coupling with ethynyl-nucleobases. A dry THF/NEt_3 or DMF/NEt_3 (4:1) solvent mixture (indicated in each case) was subjected to deoxygenation by three freeze-pump-thaw cycles with argon. Then, this solvent was added over the system containing the corresponding ethynyl-substituted base (1 eq), excess of the iodoarene derivative (indicated in each case), CuI (0.01 eq) and $Pd(PPh_3)_4$ (0.02 eq). The mixture was stirred under argon at 40 °C (unless indicated otherwise) for a period of time (indicated in each case) until completion, which was monitored by TLC. The purification methods are indicated in each case.

Mono-coupling products.

Cytosine mono-coupling products.

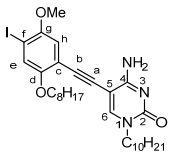


C1$_{GC1}$. **C1$_{GC1}$** was prepared following the *Standard Procedure F* for Sonogashira couplings. **C2$_{Alk10}$** (1.1 g, 3.9 mmol), diiodobenzene **B1** (2.6 g, 7.8 mmol), CuI (7.4 mg, 0.04 mmol) and $Pd(PPh_3)_4$ (90.1 mg, 0.08 mmol) were dissolved in the THF/NEt_3 mixture (24 mL). The reaction was completed in 8 h. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **C1$_{GC1}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording **C1$_{GC1}$** as a white solid (1.66 g, 89%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.14 (s, 1H, H^6), 7.77 (d, J = 8.4 Hz, 2H, H^e), 7.60 (bs, 1H, NH-H), 7.37 (d, J = 8.4 Hz, 2H, H^d), 6.94 (bs, 1H, NH-H), 3.68 (t, J = 7.2 Hz , 2H, $N^1CH_2C_9H_{19}$), 1.63-1.51 (m, 2H, $N^1CH_2CH_2C_8H_{17}$), 1.32-1.15 (m, 14H, $N^1C_2H_2C_7H_{14}CH_3$), 0.83 (t, J = 6.9 Hz, 3H, $N^1C_9H_{18}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 164.0, 154.0, 150.1, 137.3, 132.8, 122.1, 94.7, 93.0, 88.3, 83.1, 49.0, 31.3, 28.9, 28.6, 28.5, 25.8, 22.1, 13.9.

HRMS (ESI+): Calculated for $C_{22}H_{29}IN_3O$ $[M+H]^+$: 478.1349. Found: 478.1350.



C1$_{GC2}$. **C1$_{GC2}$** was prepared following the *Standard Procedure F* using **C2$_{Alk10}$** (0.77 g, 2.80 mmol), diiodoarene **B2** (3.42 g, 7.01 mmol), CuI (5.3 mg, 0.03 mmol) and $Pd(PPh_3)_4$ (64.7 mg, 0.06 mmol) in THF/NEt_3 (10 mL). The reaction was completed in 5 h. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **C1$_{GC2}$** was separated from its regioisomer **C1'$_{GC2}$** and purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). Product **C1$_{GC2}$** was obtained as a white solid (1.09 g, 61%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.54 (s, 1H, H^6), 7.31 (s, 1H, H^e), 6.81 (s, 1H, H^h), 6.13 (bs, 2H, NH_2), 3.99 (t, J = 6.7 Hz, 2H, $OCH_2C_7H_{15}$), 3.84 (s, 3H, OCH_3), 3.81 (t, J = 7.4 Hz, 2H, $N^1CH_2C_9H_{19}$), 1.84-1.66 (m,

4H, $N^1CH_2CH_2C_8H_{17}$, $OCH_2CH_2C_6H_{13}$), 1.49-1.40 (m, 2H, $OC_2H_4CH_2C_7H_{15}$), 1.38-1.17 (m, 22H, $N^1C_2H_2C_7H_{14}CH_3$, $OC_3H_6C_4H_8CH_3$), 0.83 (m, 6H, $N^1C_9H_{18}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 154.0, 155.3, 147.0, 132.2, 132.0, 131.9, 128.6, 128.4, 123.1, 113.3, 112.1, 110.9, 77.2, 69.6, 67.2, 57.0, 50.8, 50.5, 31.9, 31.8, 29.7, 29.50, 29.47, 29.34, 29.29, 29.25, 26.6, 25.9, 22.7, 14.1.

HRMS (ESI+): Calculated for $C_{31}H_{47}IN_3O_3$ $[M+H]^+$: 636.2656. Found: 636.2662.

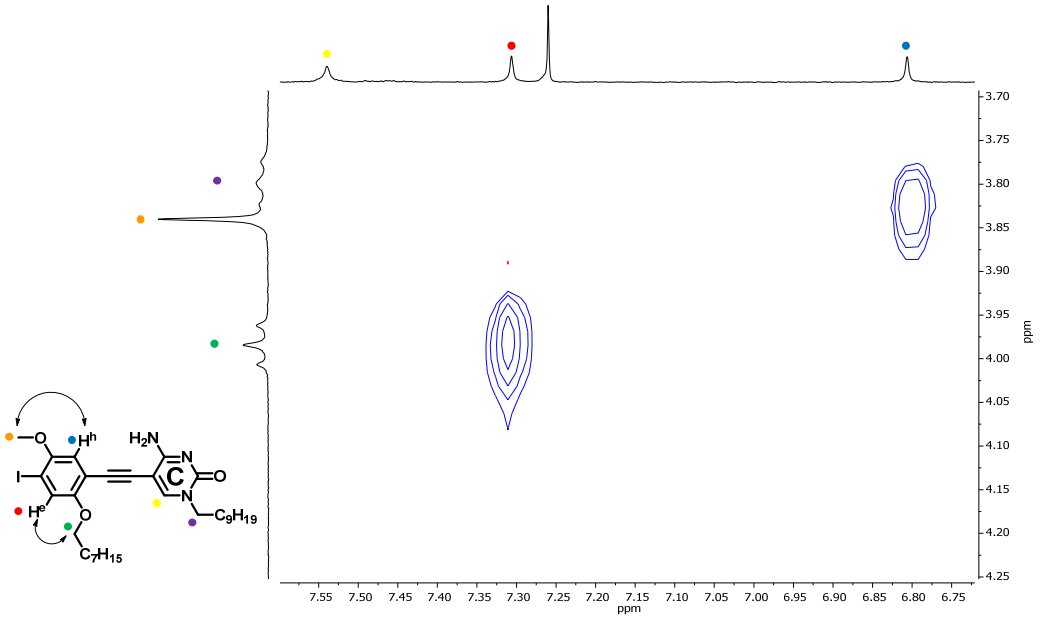
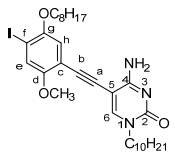


Figure 79. Region of the NOESY NMR spectrum of **C1$_{GC2}$** in $CDCl_3$ showing crosspeaks between the H^e (*ortho* to the –I substituent) and octyloxy CH_2; and H^h (*ortho* to the alkynyl substituent) and methoxy group proton signals.



C1'$_{GC2}$. Regioisomer **C1'$_{GC2}$** was obtained in the reaction for the synthesis of **C1$_{GC2}$** and was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). Product **C1'$_{GC2}$** was obtained as a white solid (0.68 g, 24%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.29 (bs, 1H, NH-*H*), 7.51 (s, 1H, H^6), 7.29 (s, 1H, H^e), 6.79 (s, 1H, H^h), 6.28 (bs, 1H, NH-*H*), 3.98-3.91 (m, 2H, $OCH_2C_7H_{15}$), 3.90 (s, 3H, OCH_3), 3.79-3.74 (m, 2H, $N^1CH_2C_9H_{19}$), 1.86-1.76 (m, 2H, $OCH_2CH_2C_6H_{13}$), 1.75-1.65 (m, 2H, $N^1CH_2CH_2C_8H_{17}$), 1.55-1.44 (m, 2H, $OC_2H_4CH_2C_7H_{15}$), 1.38-1.19 (m, 22H, $N^1C_2H_2C_7H_{14}CH_3$, $OC_3H_6C_4H_8CH_3$), 0.91-0.82 (m, 6H, $N^1C_9H_{18}CH_3$, $OC_7H_{14}CH_3$).

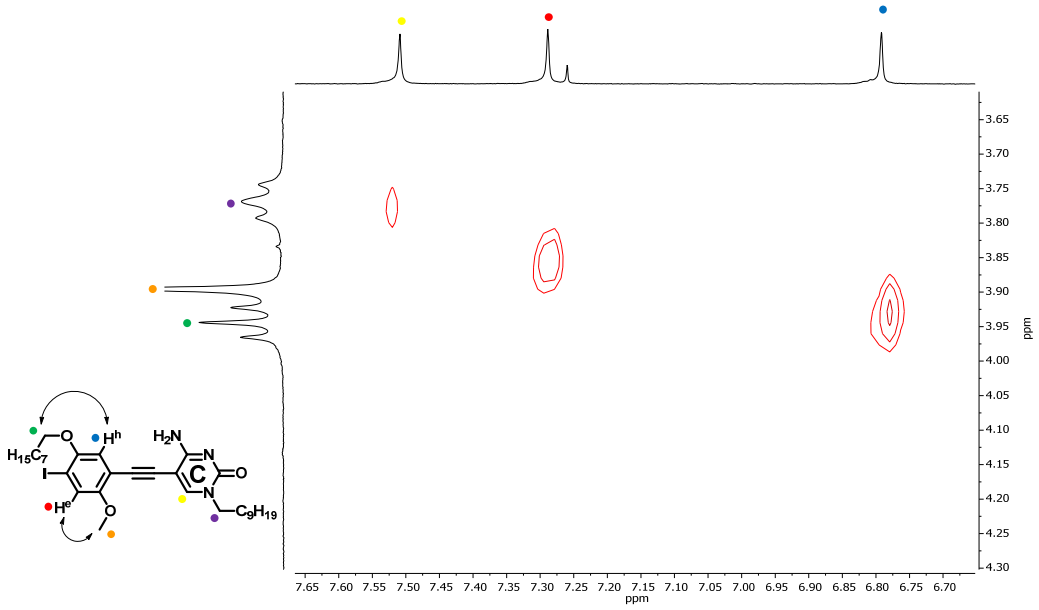
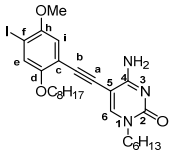


Figure 80. Region of the NOESY NMR spectrum of **C1'$_{GC2}$** in $CDCl_3$ showing crosspeaks between the H^e (*ortho* to the –I substituent) and methoxy group; and H^h (*ortho* to the alkynyl substituent) and octyloxy CH_2 proton signals.



C1$_{GC3}$. **C1$_{GC3}$** was prepared following the *Standard Procedure F* using **C2$_{Alk6}$** (0.27 g, 1.2 mmol), diiodoarene **B2** (1.5 g, 3.1 mmol), CuI (2.4 mg, 0.01 mmol) and $Pd(PPh_3)_4$ (28.9 mg, 0.02 mmol) in THF/NEt_3 (10 mL). The reaction was completed in 6 h. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **C1$_{GC3}$** was separated from its regioisomer and purified and by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). Product **C1$_{GC3}$** was obtained as a white solid (0.22 g, 30%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.54 (s, 1H, H^6), 7.31 (s, 1H, H^e), 6.81 (s, 1H, H^h), 6.07 (bs, 2H, NH_2), 3.99 (t, J = 6.7 Hz, 2H, $OCH_2C_7H_{15}$), 3.84 (s, 3H, OCH_3), 3.81 (t, J = 7.4 Hz, 2H, $N^1CH_2C_5H_{11}$), 1.84-1.68 (m, 4H, $N^1CH_2CH_2C_4H_9$, $OCH_2CH_2C_6H_{13}$), 1.50-1.39 (m, 2H, $OC_2H_4CH_2C_5H_{11}$), 1.37-1.23 (m, 14H, $N^1C_2H_2C_3H_6CH_3$, $OC_3H_6C_4H_8CH_3$), 0.92-0.84 (m, 6H, $N^1C_5H_{10}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.3, 154.9, 154.0, 152.3, 146.9, 123.1, 113.2, 90.5, 87.0, 85.5, 77.2, 69.6, 57.0, 50.8, 31.8, 31.4, 30.9, 29.33, 29.29, 29.25, 29.2, 26.3, 25.9, 22.6, 22.5, 14.1, 14.0.

HRMS (ESI+): Calculated for $C_{27}H_{39}IN_3O_3$ $[M+H]^+$: 580.2030. Found: 580.2010.

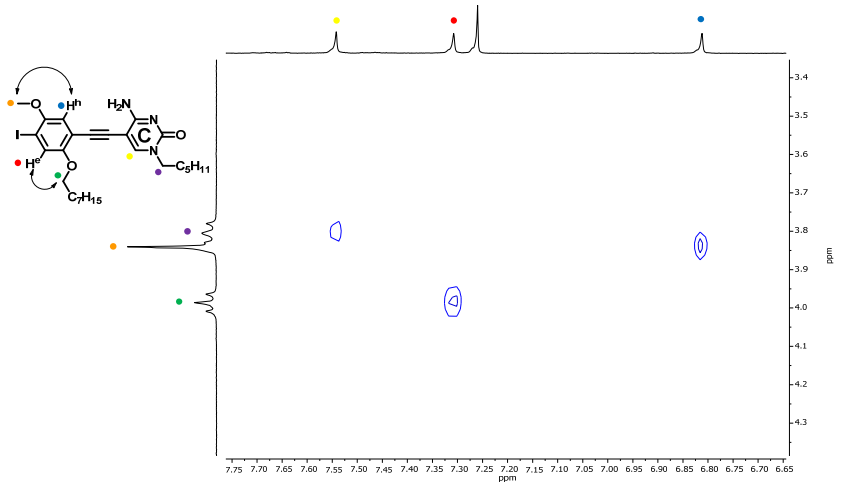
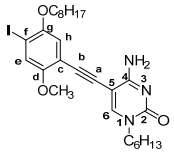


Figure 81. Region of the NOESY NMR spectrum of **C1$_{GC3}$** in $CDCl_3$ showing crosspeaks between the H^e (*ortho* to the –I substituent) and octyloxy CH_2; and H^h (*ortho* to the alkynyl substituent) and methoxy group proton signals.



C1'$_{GC3}$. Regioisomer **C1'$_{GC3}$** was obtained in the reaction for the synthesis of **C1$_{GC3}$** and was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). Product **C1'$_{GC3}$** was obtained as a white solid (0.31 g, 42%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.55 (s, 1H, H^6), 7.31 (s, 1H, H^e), 6.80 (s, 1H, H^h), 6.28 (bs, 1H, NH-*H*), 6.14 (bs, 1H, NH-*H*), 3.97-3.93 (m, 2H, $OCH_2C_7H_{15}$), 3.87 (s, 3H, OCH_3), 3.83-3.78 (m, 2H, $N^1CH_2C_5H_{11}$), 1.87-1.77 (m, 2H, $OCH_2CH_2C_6H_{13}$), 1.76-1.67 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.56-1.45 (m, 2H, $OC_2H_4CH_2C_5H_{11}$), 1.40-1.22 (m, 14H, $N^1C_2H_2C_3H_6CH_3$, $OC_3H_6C_4H_8CH_3$), 0.93-0.83 (m, 6H, $N^1C_5H_{10}CH_3$, $OC_7H_{14}CH_3$).

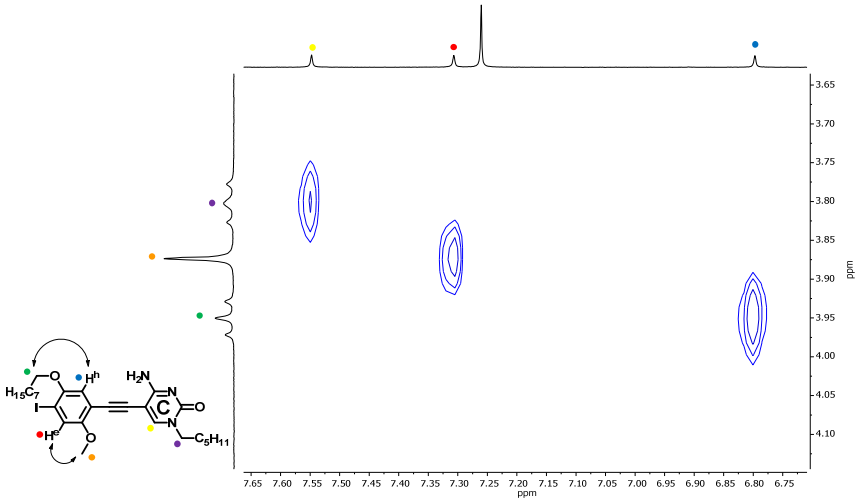
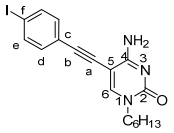


Figure 82. Region of the NOESY NMR spectrum of **C1'$_{GC3}$** in $CDCl_3$ showing crosspeaks between the H^e (*ortho* to the –I substituent) and methoxy group; and H^h (*ortho* to the alkynyl substituent) and octyloxy CH_2 proton signals.

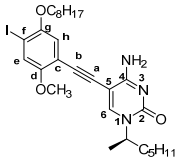


C1$_{GC4}$. **C1$_{GC4}$** was prepared following the *Standard Procedure F* using **C2$_{Alk6}$** (0.10 g, 0.46 mmol), diiodobenzene **B1** (0.45 g, 1.4 mmol), CuI (0.9 mg, 0.005 mmol) and $Pd(PPh_3)_4$ (10.5 mg, 0.01 mmol) in THF/NEt_3 (6 mL). The reaction was completed in 6 h. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **C1$_{GC4}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). Product **C1$_{GC4}$** was obtained as a white solid (147 mg, 77%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 9.28 (bs, 1H, NH_2), 7.62 (d, J = 8.4 Hz, 2H, H^e), 7.48 (s, 1H, H^6), 7.13 (d, J = 8.4 Hz, 2H, H^d), 5.94 (bs, 2H, NH_2), 3.73-3.68 (m, 2H, $N^1CH_2C_5H_{11}$), 1.71-1.58 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.31-1.20 (m, 6H, $N^1C_2H_2C_3H_6CH_3$), 0.82 (m, 3H, $N^1C_5H_{10}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.6, 154.7, 147.8, 137.4, 132.6, 121.6, 94.5, 94.3, 90.5, 81.3, 50.4, 31.2, 29.0, 26.0, 22.3, 13.8.

HRMS (ESI+): Calculated for $C_{18}H_{21}IN_3O$ $[M+H]^+$: 422.0723. Found: 422.0723.



C1$_{GC5}$. **C1$_{GC5}$** was prepared following the *Standard Procedure F* using the **C1$_{Alk6c}$** crude (88.2 mg, 0.38 mmol), diiodoarene **B2** (0.55 g, 1.13 mmol), CuI (0.7 mg, 0.004 mmol) and $Pd(PPh_3)_4$ (8.7 mg, 0.008 mmol) in the THF/NEt_3 mixture (5 mL). The reaction was completed in 8 h. After solvent evaporation, the crude was separated from its regioisomer and purified by column chromatography eluted with $CHCl_3$/MeOH (80:1). **C1$_{GC5}$** was obtained as a white solid (18.1 mg, 9%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.63 (bs, 1H, NH-H), 7.52 (s, 1H, H^6), 7.30 (s, 1H, H^e), 6.81 (s, 1H, H^h), 6.21 (bs, 1H, NH-H), 4.89 (m, 1H, $CH(C_5H_{11})CH_3$), 3.95 (t, J = 6.7 Hz, 2H, $OCH_2C_7H_{15}$), 3.90 (s, 3H, OCH_3), 1.82 (m, 3H, $N^1CH(C_5H_{11})CH_3$), 1.62 (m, 2H, $N^1CH(CH_3)CH_2C_4H_9$), 1.51 (m, 2H, $OCH_2CH_2C_6H_{13}$), 1.4-1.2 (m, 16H, $N^1CH(CH_3)CH_2C_3H_6CH_3$, $OC_2H_4C_5H_{10}CH_3$), 1.0-0.8 (m, 6H, $N^1C_5H_{10}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 163.9, 155.0, 154.3, 151.9, 143.1, 121.7, 114.6, 111.9, 92.3, 91.1, 87.9, 86.0, 70.2, 56.5, 52.0, 36.0, 31.8, 31.5, 29.7, 29.24, 29.21, 29.1, 26.1, 25.7, 22.7, 22.4, 20.4, 14.1, 14.0.

HRMS (MALDI, dithranol): Calculated for $C_{28}H_{41}IN_3O_3$ $[M+H]^+$: 594.2187. Found: 594.2188.

$[\alpha]_D^{20}$ = + 24.1 (c = 5.4 x 10^{-4} g/mL, CH_2Cl_2)

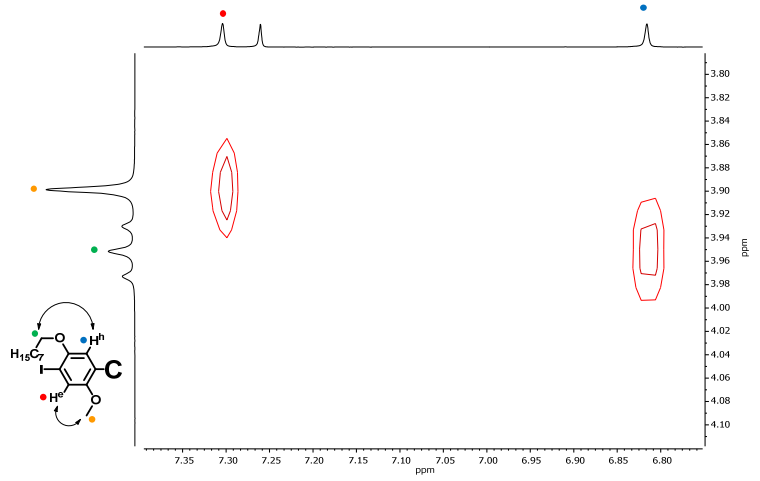
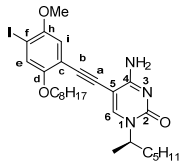
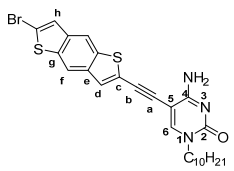


Figure 83. Region of the NOESY NMR spectrum of **C1$_{GC5}$** in $CDCl_3$ showing crosspeaks between the H^e (*ortho* to the –I substituent) and methoxy group; and H^h (*ortho* to the alkynyl substituent) and octyloxy CH_2 proton signals.



C1'$_{GC5}$. Regioisomer **C1'$_{GC5}$** was obtained in the reaction for the synthesis of **C1$_{GC5}$** and was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (80:1). Product **C1'$_{GC5}$** was obtained as a white solid (6.1 mg, 3%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.54 (s, 1H, H^6), 7.31 (s, 1H, H^e), 6.83 (s, 1H, H^h), 6.23 (bs, 1H, NH-*H*), 6.11 (bs, 1H, NH-*H*), 4.92 (m, 1H, C*H*(C_5H_{11})), 3.99 (t, *J* = 6.7 Hz, 2H, $OCH_2C_7H_{15}$), 3.84 (s, 3H, OCH_3), 1.79 (m, 3H, N^1CHCH_3), 1.63 (m, 2H, $N^1CHCH_2C_4H_9$), 1.5-1.2 (m, 18H, N^1 $CHCH_2C_3H_6CH_3$, $OCH_2C_6H_{12}CH_3$), 1.0-0.8 (m, 6H, $N^1C_5H_{10}CH_3$, $OC_7H_{14}CH_3$).

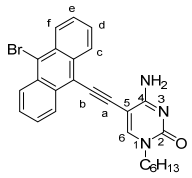


C1$_{GC8}$. **C1$_{GC8}$** was prepared following the *Standard Procedure F* with **C1$_{Alk10}$** (0.34 g, 1.23 mmol), dibromobenzodithiophene **B3** (1.3 g, 3.7 mmol), CuI (2.3 mg, 0.01 mmol) and $Pd(PPh_3)_4$ (28 mg, 0.02 mmol) in the DMF/NEt_3 mixture (35 mL). The reaction was completed overnight. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **C1$_{GC8}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (30:1), affording **C1$_{GC8}$** as an ochre solid (440 mg, 67%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.11 (s, 1H, H^f), 8.08 (s, 1H, H^f), 7.64 (s, 1H, H^6), 7.48 (s, 1H, H^h), 7.36 (s, 1H, H^d), 6.78 (bs, 1H, NH-*H*), 5.67 (bs, 1H, NH-*H*), 3.84-3.79 (m, 2H, $N^1CH_2C_5H_{11}$), 1.80-1.69 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.40-1.21 (m, 6H, $N^1C_2H_2C_3H_6CH_3$), 0.92-0.83 (m, 3H, $N^1C_5H_{10}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.0, 154.7, 148.8, 138.7, 138.2, 137.8, 136.4, 128.3, 125.7, 122.6, 116.9, 116.4, 115.5, 89.7, 88.9, 86.0, 50.9, 31.9, 29.50, 29.46, 29.3, 29.2, 26.6, 22.7, 14.1.

HRMS (ESI+): Calculated for $C_{26}H_{29}N_3OS_2Br$ $[M+H]^+$: 542.0929. Found: 542.0933.

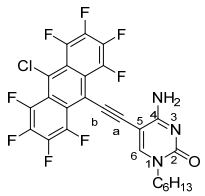


C1$_{GC9}$. **C1$_{GC9}$** was prepared following the *Standard Procedure F* using **C1$_{Alk6}$** (170 mg, 0.76 mmol), dibromoanthracene **B4** (1.8 g, 5.4 mmol), CuI (1.5 mg, 0.01 mmol) and $Pd(PPh_3)_4$ (18 mg, 0.02 mmol), dissolved in the THF/NEt_3 mixture (25 mL). The reaction mixture was heated at 60 °C overnight. Once completed, the mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **C1$_{GC9}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1). **C1$_{GC9}$** was obtained as a brown solid (173 mg, 47%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.59-8.55 (m, 2H, H^f), 8.51-8.48 (m, 2H, H^e), 7.92 (bs, 1H, NH-*H*), 7.72 (s, 1H, H^6), 7.67-7.60 (m, 4H, H^c, H^d), 5.99 (bs, 1H, NH-*H*), 3.83-3.78 (m, 2H, $N^1CH_2C_5H_{11}$), 1.83-1.69 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.40-1.26 (m, 6H, $N^1C_2H_2C_3H_6CH_3$), 0.91-0.86 (m, 3H, $N^1C_5H_{10}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 163.1, 151.5, 132.1, 129.5, 128.6, 128.5, 127.5, 127.4, 123.1, 117.8, 93.6, 89.8, 49.3, 30.9, 28.6, 25.6, 22.0, 13.9.

HRMS (ESI+): Calculated for $C_{26}H_{25}N_3OBr$ $[M+H]^+$: 474.1175. Found: 474.1190.



C1$_{GC10}$. **C1$_{GC10}$** was prepared following the *Standard Procedure F* using **C1$_{Alk6}$** (127 mg, 0.58 mmol), 9,10-dichlorooctafluoroanthracene **B5** (272 mg, 0.69 mmol), CuI (1 mg, 0.01 mmol) and $Pd(PPh_3)_4$ (4 mg, 0.01 mmol) in the DMF/NEt_3 mixture (20 mL) and heating at 50 °C overnight. Once completed, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **C1$_{GC10}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording **C1$_{GC10}$** as a brown solid (43 mg, 13%).

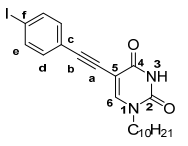
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.24 (s, 1H, H^6), 7.84 (bs, 1H, NH-*H*), 6.49 (bs, 1H, NH-*H*), 3.84-3.74 (m, 2H, $N^1CH_2C_5H_{11}$), 1.69-1.56 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.38-1.22 (m, 6H, $N^1C_2H_2C_3H_6CH_3$), 0.93-0.83 (m, 3H, $N^1C_5H_{10}CH_3$).

^{19}F NMR (300 MHz, DMSO-d_6) δ(ppm) = -136.5 (dd, 2F), -138.5 (dd, 2F), -152.6 (dd, 2F), -153.2 (dd, 2F).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 133.0, 132.22, 132.17, 132.15, 131.6, 131.5, 129.6, 128.9, 128.8, 118.1, 79.2, 79.0, 78.7, 45.8, 30.9, 29.1, 25.6, 22.1, 13.9.

HRMS (ESI+): Calculated for $C_{26}H_{17}ClF_8N_3O$ $[M+H]^+$: 574.0927. Found: 574.0925.

Uracil mono-coupling products.

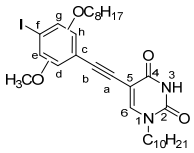


U1$_{AU1}$. **U1$_{AU1}$** was prepared following the *Standard Procedure F* with **U1$_{Alk10}$** (0.44 g, 1.6 mmol), diiodobenzene **B1** (2.6 g, 8.0 mmol), CuI (3.0 mg, 0.02 mmol) and $Pd(PPh_3)_4$ (3.7 mg, 0.03 mmol) in the THF/NEt_3 mixture (30 mL). The reaction was completed in 6 h. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **U1$_{AU1}$** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (80:1), affording **U1$_{AU1}$** as a white solid (0.59 g, 78%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.63 (bs, 1H, N*H*), 8.21 (s, 1H, H^6), 7.77 (d, J = 8.0 Hz, 2H, H^e), 7.23 (d, J = 8.4 Hz, 2H, H^d), 3.68 (t, J = 7.0 Hz , 2H, $N^1CH_2C_9H_{19}$), 1.65-1.52 (m, 2H, $N^1CH_2CH_2C_8H_{17}$), 1.31-1.15 (m, 14H, $N^1C_2H_2C_7H_{14}CH_3$), 0.84 (t, J = 7.0 Hz, 3H, $N^1C_9H_{18}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 161.6, 149.7, 146.8, 137.5, 133.0, 122.1, 99.7, 94.5, 92.9, 81.5, 49.4, 31.8, 29.43, 29.37, 29.2, 29.1, 26.3, 22.6, 14.1.

HRMS (ESI+): Calculated for $C_{22}H_{28}IN_2O_2$ $[M+H]^+$: 479.1190. Found: 479.1181.

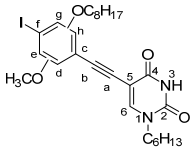


U1$_{AU2}$. **U1$_{AU2}$** was prepared following the *Standard Procedure F* with **U1$_{Alk10}$** (0.14 g, 0.5 mmol), diiodoarene **B2** (0.62 g, 1.3 mmol), CuI (1.0 mg, 0.005 mmol) and $Pd(PPh_3)_4$ (11.7 mg, 0.01 mmol) in the THF/NEt_3 mixture (6 mL). The reaction was completed in 6 h. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). **U1$_{AU2}$** was obtained as a white solid corresponding to the mixture of the two regioisomers (0.19 g, 59%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.07 (bs, 2H, N*H*), 7.54 (s, 1H, H^6), 7.48 (s, 1H, H^6), 7.29 (s, 1H, H^e), 7.28 (s, 1H, H^e), 6.91 (s, 2H, H^d), 3.97 (t, J = 7.0 Hz , 2H, $OCH_2C_7H_{15}$), 3.94 (t, J = 7.0 Hz , 2H, $OCH_2C_7H_{15}$), 3.85 (s, 3H, OCH_3), 3.83 (s, 3H, OCH_3), 3.75 (t, J = 7.0 Hz , 4H, $N^1CH_2C_9H_{19}$), 1.80 (m, 4H, $OCH_2CH_2C_6H_{13}$), 1.71 (m, 4H, $N^1CH_2CH_2C_8H_{17}$), 1.49 (m, 4H, $OC_2H_4CH_2C_5H_{11}$), 1.38-1.19 (m, 44H, $N^1C_2H_2C_7H_{14}CH_3$, $OC_3H_6C_4H_8CH_3$), 0.93-0.83 (m, 12H, $N^1C_9H_{18}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 161.8, 161.7, 154.4, 154.2, 152.2, 151.8, 149.8, 146.7, 146.5, 124.0, 122.1, 116.3, 114.7, 113.1, 112.3, 100.1, 99.9, 89.9, 89.8, 88.0, 87.0, 84.81, 84.76, 70.1, 69.9, 56.9, 56.6, 49.3, 31.8, 31.7, 29.6, 29.4, 29.3, 29.24, 29.16, 29.1, 29.0, 26.3, 26.0, 25.9, 22.6, 14.0.

HRMS (ESI+): Calculated for $C_{31}H_{46}IN_2O_4$ $[M+H]^+$: 637.2496. Found: 637.2524.



U1$_{AU3}$. **U1$_{AU3}$** was prepared following the *Standard Procedure F*. **U1$_{Alk6}$** (0.09 g, 0.4 mmol), diiodoarene **B2** (0.85 g, 1.7 mmol), CuI (0.8 mg, 0.004 mmol) and $Pd(PPh_3)_4$ (9.4 mg, 0.008 mmol) were dissolved in the THF/NEt_3 mixture (5 mL). The reaction was completed in 8 h. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. The crude was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1). **U1$_{AU3}$** was obtained as a white solid corresponding to the mixture of the two regioisomers (0.22 g, 93%).

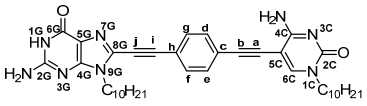
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.15 (bs, 2H, N*H*), 7.54 (s, 1H, H^6), 7.48 (s, 1H, H^6), 7.29 (s, 1H, H^e), 7.28 (s, 1H, H^e), 6.91 (s, 2H, H^d), 3.99-3.92 (m, 4H, $OCH_2C_7H_{15}$), 3.85 (s, 3H, OCH_3), 3.83 (s, 3H, OCH_3), 3.75 (t, J = 7.0 Hz , 4H, $N^1CH_2C_5H_{11}$), 1.81 (m, 4H, $OCH_2CH_2C_6H_{13}$), 1.72 (m, 4H, $N^1CH_2CH_2C_4H_9$), 1.49 (m, 4H, $OC_2H_4CH_2C_5H_{11}$), 1.41-1.23 (m, 28H, $N^1C_2H_2C_3H_6CH_3$, $OC_3H_6C_4H_8CH_3$), 0.97-0.84 (m, 12H, $N^1C_9H_{18}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 161.9, 161.8, 154.2, 154.0, 152.1, 151.7, 149.8, 146.7, 146.5, 123.9, 122.0, 116.1, 114.6, 113.0, 112.3, 100.0, 99.7, 89.6, 87.9, 86.8, 84.9, 70.0, 69.8, 56.8, 56.5, 49.2, 31.7, 31.6, 31.1, 29.14, 29.09, 29.0, 28.9, 25.88, 25.85, 25.8, 22.5, 22.3, 14.0, 13.9, 13.8.

HRMS (ESI+): Calculated for $C_{27}H_{38}IN_2O_4$ $[M+H]^+$: 581.1870. Found: 581.1875.

Unsymmetric monomers.

GC monomers.

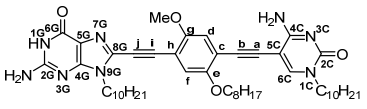


GC1. **GC1** was prepared following the *Standard Procedure F* using $\mathbf{G1_{Alk10}}$ (0.49 mg, 1.55 mmol), $\mathbf{C1_{GC1}}$ (0.74 g, 1.55 mmol), CuI (3.0 mg, 0.02 mmol) and $Pd(PPh_3)_4$ (35.9 mg, 0.03 mmol) suspended in the THF/NEt_3 mixture (10 mL). The reaction was completed overnight. Then, the resulting insoluble yellow solid in suspension was filtered and washed with THF. **GC1** was precipitated by adding NEt_3 to a solution of the obtained solid dissolved in $CHCl_3$ with a few drops of TFA. The resulting solid was washed with acetonitrile, affording **GC1** as an ochre solid (0.43 g, 42%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 9.56 (bs, 1H, $N^{1G}H$), 7.90 (s, 1H, H^{6C}), 7.66 (dd, J = 8.8, J' = 3.4 Hz, 4H, H^d, H^e, H^f, H^g), 7.51 (bs, 2H, $C^{4C}NH_2$), 4.32 (t, J = 7.2 Hz, 2H, $N^{8G}CH_2C_9H_{19}$), 3.93 (t, J = 7.2 Hz, 2H, $N^{1C}CH_2C_9H_{19}$), 1.97-1.87 (m, 2H, $N^{8G}CH_2CH_2C_8H_{17}$), 1.84-1.72 (m, 2H, $N^{1C}CH_2CH_2C_8H_{17}$), 1.47-1.20 (m, 28H, $N^{8G}C_2H_4C_7H_{14}CH_3$, $N^{1C}C_2H_4C_7H_{14}CH_3$), 0.90-0.82 (m, 6H, $N^{8G}C_9H_{18}CH_3$, $N^{1C}C_9H_{18}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 159.2, 155.3, 151.0, 149.5, 146.6, 132.9, 132.3, 130.2, 124.7, 120.2, 118.7, 116.5, 112.7, 108.9, 107.7, 106.5, 97.8, 91.9, 77.9, 72.1, 51.7, 46.3, 31.81, 31.76, 29.4, 29.3, 29.19, 29.15, 29.0, 28.93, 28.87, 28.7, 26.3, 26.2, 22.60, 22.56, 13.9, 13.8.

HRMS (ESI+): Calculated for $C_{39}H_{53}N_8O_2$ $[M+H]^+$: 665.4286. Found: 665.4274.

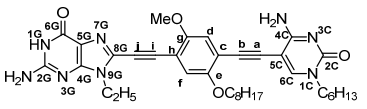


GC2. **GC2** was prepared following the *Standard Procedure F*. $\mathbf{G1_{Alk10}}$ (168.4 mg, 0.53 mmol), $\mathbf{C1_{GC2}}$ (339.4 mg, 0.53 mmol), CuI (1.0 mg, 0.005 mmol) and $Pd(PPh_3)_4$ (12.4 mg, 0.01 mmol) were suspended in the THF/NEt_3 mixture (10 mL). The reaction was completed overnight. Then, the resulting insoluble bright yellow solid in suspension was filtered and washed with THF. **GC2** was precipitated by adding NEt_3 to a solution of the obtained solid dissolved in $CHCl_3$ with a few drops of TFA. The resulting solid was washed with acetonitrile, affording **GC2** as a bright yellow solid (0.28 g, 64%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 9.86 (bs, 1H, $N^{1G}H$), 7.83 (s, 1H, H^{6C}); 7.48 (bs, 2H, $C^{4C}NH_2$), 7.10 (s, 1H, H^f), 7.03 (s, 1H, H^d), 4.35 (t, J = 7.5 Hz, 2H, $N^{9G}CH_2C_9H_{19}$), 4.08 (t, J = 6.8 Hz, 2H, $OCH_2C_7H_{15}$), 3.96-3.88 (m, 5H, $N^{1C}CH_2C_9H_{19}$, OCH_3), 1.98-1.86 (m, 2H, $OCH_2CH_2C_6H_{13}$), 1.86-1.70 (m, 4H, $N^{9G}CH_2CH_2C_8H_{17}$, $N^{1C}CH_2CH_2C_8H_{17}$), 1.48-1.16 (m, 38H, $N^{9G}C_2H_2C_7H_{14}CH_3$, $N^{1C}C_2H_2C_7H_{14}CH_3$, $OC_2H_4C_5H_{10}CH_3$), 0.92-0.81 (m, 9H, $N^{9G}C_9H_{18}CH_3$, $N^{1C}C_9H_{18}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 159.6, 156.1, 155.2, 153.5, 149.4, 149.3, 146.3, 130.4, 120.2, 116.5, 116.3, 115.7, 114.2, 112.7, 109.0, 108.9, 107.5, 104.3, 95.9, 92.0, 82.6, 69.9, 56.3, 51.6, 46.1, 31.78, 31.75, 31.6, 31.4, 29.4, 29.33, 29.26, 29.2, 29.1, 29.0, 28.93, 28.89, 28.8, 28.7, 26.4, 26.2, 25.6, 22.59, 22.55, 22.5, 13.93, 13.88, 13.8.

MS (MALDI, dithranol+NaI): Calculated for $C_{48}H_{71}N_8O_4$ $[M+H]^+$: 823.5593. Found: 823.5629.



GC3. **GC3** was prepared following the *Standard Procedure F* with $\mathbf{G1_{Alk2}}$ (50.0 mg, 0.25 mmol), $\mathbf{C1_{GC3}}$ (171.1 mg, 0.30 mmol), CuI (0.5 mg, 0.002 mmol) and $Pd(PPh_3)_4$ (56.9 mg, 0.049 mmol) suspended in the DMF/NEt_3 mixture (4 mL). The reaction was completed overnight. The resulting insoluble solid in suspension was filtered and washed with THF. The solid was dried under reduced pressure, affording **GC3** as a bright yellow solid (83.9 mg, 52%).

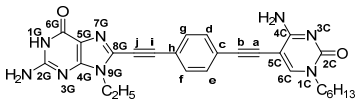
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.81 (s, 1H, H^{6C}), 7.38 (bs, 2H, $C^{4C}NH_2$), 7.10 (s, 1H, H^f), 7.01 (s, 1H, H^d), 4.45-4.38 (m, 2H, $N^{9G}CH_2CH_3$), 4.07 (t, J = 6.8 Hz, 2H, $OCH_2C_7H_{15}$), 3.94-3.87 (m, 5H, $N^{1C}CH_2C_5H_{11}$,

OCH_3), 1.84-1.73 (m, 4H, $N^{1C}CH_2CH_2C_4H_9$, $OCH_2CH_2C_6H_{13}$), 1.57 (t, *J* = 6.6 Hz, 3H, $N^{9G}CH_2CH_3$), 1.46-1.21 (m, 16H, $N^{1C}C_2H_2C_3H_6CH_3$, $OC_2H_4C_5H_{10}CH_3$), 0.95-0.82 (m, 6H, $N^{1C}C_5H_{10}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 156.1, 154.4, 154.1, 153.1, 152.3, 151.0, 147.1, 129.3, 120.9, 117.1, 116.6, 115.6, 113.5, 113.3, 111.3, 109.5, 92.5, 90.4, 90.0, 84.3, 83.7, 69.3, 56.5, 49.6, 31.3, 30.9, 28.8, 28.6, 28.1, 25.6, 25.4, 22.2, 22.0, 14.8, 14.0, 13.9.

FT-IR (KBr), *v* (cm^{-1}): 3404, 3325, 3150, 2953, 2930, 2856, 1688, 1649, 1609, 1560, 1504, 1485, 1466, 1412, 1389, 1317, 1281, 1221, 1198, 1124, 1076, 1034, 854, 781, 714, 662, 482, 530.

HRMS (ESI+): Calculated for $C_{36}H_{47}N_8O_4$ $[M+H]^+$: 655.3714. Found: 655.3727.

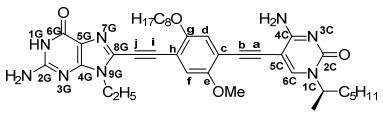


GC4. **GC4** was prepared following the *Standard Procedure F* with **G1$_{Alk2}$** (35.4 mg, 0.17 mmol), **C1$_{GC4}$** (147.0 mg, 0.35 mmol), CuI (0.3 mg, 0.002 mmol) and $Pd(PPh_3)_4$ (4.0 mg, 0.004 mmol) suspended in the THF/NEt_3 mixture (5 mL). The reaction was completed in 6 h. Then, the mixture was concentrated under reduced pressure. **GC4** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (10:1), affording **GC4** as an ochre solid (66.2 mg, 76%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.86 (bs, 1H, $N^{1G}H$), 8.22 (s, 1H, H^{6C}), 7.75 (bs, 1H, $C^{4C}NH$-*H*), 7.68 (s, 4H, H^d, H^e, H^f, H^g), 7.14 (bs, 1H, $C^{4C}NH$-*H*), 6.71 (bs, 2H, $C^{2G}NH_2$), 4.11 (q, *J* = 7.2 Hz, 2H, $N^{9G}CH_2CH_3$), 3.71 (t, *J* = 7.2 Hz, 2H, $N^{1C}CH_2C_5H_{11}$), 1.59 (m, 2H, $N^{1C}CH_2CH_2C_4H_9$), 1.35 (t, *J* = 7.2 Hz, 3H, $N^{9G}CH_2CH_3$), 1.31-1.19 (m, 6H, $N^{1C}C_2H_2C_3H_6CH_3$), 0.86 (t, 3H, $N^{1C}C_5H_{10}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 159.1, 156.0, 154.1, 153.2, 150.9, 147.0, 131.79, 131.75, 128.9, 122.8, 121.0, 116.8, 95.0, 92.7, 89.6, 80.8, 80.7, 49.6, 38.1, 30.8, 28.0, 25.3, 21.9, 14.9, 13.8.

HRMS (ESI+): Calculated for $C_{27}H_{29}N_8O_2$ $[M+H]^+$: 497.2407. Found: 497.2408.



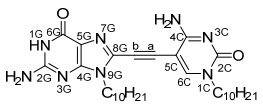
GC5. **GC5** was prepared following the *Standard Procedure F* with **G1$_{Alk2}$** (5.0 mg, 0.02 mmol), **C1$_{GC5}$** (16.2 mg, 0.03 mmol), CuI (0.04 mg, 0.0002 mmol) and $Pd(PPh_3)_4$ (0.5 mg, 0.0005 mmol) were suspended in the THF/NEt_3 mixture (5 mL) and stirred overnight. Once completed, the mixture was filtered through celite plug and washed with THF. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (10:1). A final recrystallization in acetonitrile afforded **GC5** as a pale solid (13.0 mg, 86%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 13.68 (bs, 1H, $N^{1G}H$), 8.22 (s, 1H, H^{6C}), 7.75 (bs, 1H, $C^{4C}NH$-*H*), 7.68 (s, 4H, H^d, H^e, H^f, H^g), 7.14 (bs, 1H, $C^{4C}NH$-*H*), 6.71 (bs, 2H, $C^{2G}NH_2$), 4.11 (q, *J* = 7.1 Hz, 2H, $N^{8G}CH_2CH_3$), 3.71 (t, *J* = 7.1 Hz, 2H, $N^{1C}CH_2C_5H_{11}$), 1.59 (m, 2H, $N^{1C}CH_2CH_2C_4H_9$), 1.35 (t, J = 7.1 Hz, 2H, $N^{8G}CH_2CH_3$), 1.22 (m, 6H, $N^{1C}C_2H_4C_3H_6CH_3$), 0.84 (m, 3H, $N^{1C}C_5H_{10}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 174.4, 174.0, 170.5, 164.6, 156.0, 141.6, 137.9, 123.5, 111.4, 92.7, 88.1, 84.3, 69.1, 45.9, 32.0, 31.9, 31.6, 29.7, 29.6, 29.5, 29.4, 26.2, 26.1, 25.6, 22.7, 22.4, 19.3, 15.4, 14.1, 14.0, 8.6.

HRMS (MALDI, dithranol): Calculated for $C_{37}H_{49}N_8O_4$ $[M+H]^+$: 669.3871. Found: 669.3881.

$[\alpha]_D^{20}$ +257 (*c* 4.3 x 10^{-4} g/mL, CH_2Cl_2)

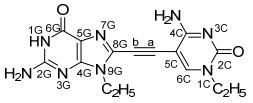


GC6. **GC6** was prepared following *Standard Procedure F*. **G1$_{Alk10}$** (0.15 g, 0.49 mmol), **C1$_{Alk10}$** (0.18 g, 0.49 mmol), CuI (1.0 mg, 0.005 mmol) and $Pd(PPh_3)_4$ (11.2 mg, 0.010 mmol) were suspended in the THF/NEt_3 mixture (7 mL). The reaction was completed overnight. The resulting insoluble dark red solid in suspension was filtered and washed with THF and acetonitrile. **GC6** was obtained as an ochre solid (77 mg, 28%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 9.58 (bs, 1H, $N^{1G}H$), 8.23 (bs, 2H, $C^{4C}NH_2$), 8.21 (s, 1H, H^{6C}), 6.80 (bs, 2H, $C^{2G}NH_2$), 4.26 (m, 2H, $N^{8G}CH_2C_9H_{19}$), 3.92 (t, 2H, $N^{1C}CH_2C_9H_{19}$), 1.84 (m, 2H, $N^{8G}CH_2CH_2C_8H_{17}$), 1.75 (m, 2H, $N^{1C}CH_2CH_2C_8H_{17}$), 1.42-1.16 (m, 28H, $N^{8G}C_2H_4C_7H_{14}CH_3$, $N^{1C}C_2H_4C_7H_{14}CH_3$), 0.91-0.82 (m, 6H, $N^{8G}C_9H_{18}CH_3$, $N^{1C}C_9H_{18}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 159.1, 155.2, 154.9, 154.6, 145.7, 129.1, 109.4, 100.0, 95.0, 91.2, 87.4, 79.6, 52.1, 47.1, 46.2, 31.8, 31.7, 29.4, 29.34, 29.25, 29.2, 29.1, 28.9, 28.6, 26.2, 26.1, 22.6, 22.5, 13.89, 13.85, 8.4.

HRMS (ESI+): Calculated for $C_{31}H_{49}N_8O_2$ $[M+H]^+$.: 565.3972. Found: 565.3889.

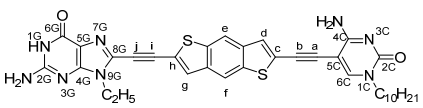


GC7. **GC7** was prepared following *Standard Procedure F*. **G1$_{Alk2}$** (40 mg, 0.20 mmol), **C1$_{Alk2}$** (0.13 g, 0.49 mmol), CuI (0.4 mg, 0.002 mmol) and $Pd(PPh_3)_4$ (4.6 mg, 0.039 mmol) were suspended in the THF/NEt_3 mixture (10 mL). The reaction was completed overnight. The resulting insoluble ochre solid in suspension was filtered and washed with THF, acetonitrile and methanol, affording **GC7** as an ochre solid (40 mg, 60%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.80 (bs, 1H, $N^{1G}H$), 8.31 (s, 1H, H^{6C}), 7.66 (bs, 1H, C^{4C}NH-H), 6.96 (bs, 1H, C^{4C}NH-H), 6.62 (bs, 2H, $C^{2G}NH_2$), 4.07 (q, J = 6.9 Hz, 2H, $N^{9G}CH_2CH_3$), 3.75 (q, J = 6.9 Hz, 2H, $N^{1C}CH_2CH_3$), 1.30 (t, J = 7.0 Hz, 3H, $N^{9G}CH_2CH_3$), 1.19 (t, J = 7.0 Hz, 3H, $N^{1C}CH_2CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 159.6, 156.3, 154.4, 154.2, 150.9, 147.5, 129.0, 117.1, 88.3, 85.0, 81.7, 45.1, 38.0, 15.0, 13.8.

HRMS (MALDI, dithranol): Calculated for $C_{15}H_{17}N_8O_2$ $[M+H]^+$: 341.1469. Found: 341.1482.

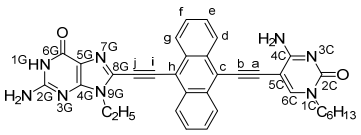


GC8. **GC8** was prepared following *Standard Procedure F*. **G1$_{Alk2}$** (20.0 mg, 0.10 mmol), **C1$_{GC8}$** (160.0 mg, 0.30 mmol), CuI (0.2 mg, 0.001 mmol) and $Pd(PPh_3)_4$ (11.4 mg, 0.010 mmol) were suspended in the DMF/NEt_3 mixture (6 mL). The mixture was stirred at 50 ^{o}C overnight. Once completed, the resulting insoluble solid in suspension was filtered and washed with THF. **GC8** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (10:1). A final recrystallization using acetonitrile yielded **GC8** as an ochre solid (31.8 mg, 49%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.91 (bs, 1H, $N^{1G}H$), 9.47 (bs, 1H, C^{4C}NH-H), 9.41 (bs, 1H, C^{4C}NH-H), 8.69 (s, 1H, H^g), 8.58 (s, 1H, H^e), 8.56 (s, 1H, H^f), 8.02 (s, 1H, H^d), 7.90 (s, 1H, H^{6C}), 4.13 (q, J = 6.9 Hz, 2H, $N^{9G}CH_2CH_3$), 3.80 (t, J = 7.2 Hz, 2H, $N^{1C}CH_2C_9H_{19}$), 1.71-1.59 (m, 2H, $N^{1C}CH_2CH_2C_8H_{17}$), 1.37 (t, J = 6.9 Hz, 3H, $N^{9G}CH_2CH_3$), 1.38-1.19 (m, 14H, $N^{1C}C_2H_4C_7H_{14}CH_3$), 0.84 (t, J = 7.2 Hz, 3H, $N^{1C}C_9H_{18}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 159.0, 158.3, 154.6, 153.6, 147.1, 138.2, 138.1, 137.4, 136.3, 129.9, 126.5, 122.0, 117.4, 116.3, 89.41, 89.36, 84.8, 84.6, 84.1, 61.6, 49.7, 31.4, 29.0, 29.0, 28.8, 28.7, 28.1, 25.7, 22.2, 14.9, 14.0.

HRMS (MALDI, dithranol): Calculated for $C_{35}H_{37}N_8O_2S_2$ $[M+H]^+$: 665.2475. Found: 665.2485.



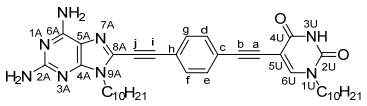
GC9. **GC9** was prepared following *Standard Procedure F*. **G1$_{Alk2}$** (40 mg, 0.20 mmol), **C1$_{GC9}$** (125 mg, 0.30 mmol), CuI (3.8 mg, 0.002 mmol) and $Pd(PPh_3)_4$ (45.5 mg, 0.039 mmol) were suspended in the DMF/NEt_3 mixture (10 mL). The reaction mixture was stirred at 50 ^{o}C overnight. Once completed, the resulting insoluble dark red solid in suspension was filtered, washed with THF and acetonitrile, affording **GC9** as a dark red solid (43 mg, 37%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.83 (bs, 1H, $N^{1G}H$), 8.89 (bs, 2H, $C^{4C}NH_2$), 8.81 (s, 1H, H^{6C}), 8.66-8.58 (m, 4H, H^d, H^e), 7.88-7.77 (m, 4H, H^f, H^g), 6.75 (bs, 2H, $C^{2G}NH_2$), 4.34 (t, J = 7.0 Hz, 2H, $N^{9G}CH_2CH_3$), 3.86 (t, J = 7.0 Hz, 2H, $N^{1C}CH_2C_5H_{11}$), 1.71 (m, 2H, $N^{1C}CH_2CH_2C_4H_9$), 1.50 (t, J = 7.0 Hz, 3H, $N^{9G}CH_2CH_3$), 1.39-1.26 (m, 8H, $N^{1C}C_2H_4C_4H_8CH_3$), 0.88 (m, 3H, $N^{1C}C_5H_{11}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 160.3, 158.1, 156.3, 154.4, 153.1, 151.2, 148.9, 131.4, 131.2, 129.2, 128.3, 127.7, 127.5, 126.0, 118.3, 117.9, 116.0, 92.6, 91.1, 89.3, 79.1, 49.5, 38.2, 30.8, 28.2, 25.4, 21.9, 15.1, 13.9.

HRMS (ESI+): Calculated for $C_{35}H_{33}N_8O_2$ $[M+H]^+$: 597.2720. Found: 597.2732.

AU monomers.

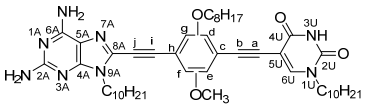


AU1. **AU1** was prepared following *Standard Procedure F* with $\mathbf{AU1_{Alk10}}$ (30.0 mg, 0.10 mmol), $\mathbf{U1_{AU1}}$ (91.3 mg, 0.19 mmol), CuI (3.0 mg, 0.001 mmol) and $Pd(PPh_3)_4$ (0.2 mg, 0.03 mmol) in the THF/NEt_3 mixture (5 mL). The reaction was completed overnight. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. **AU1** was purified by column chromatography on silica gel eluted with toluene/acetone (3:1). A final recrystallization using acetonitrile yielded **AU1** as a yellow solid (52.1 mg, 82%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.65 (bs, 1H, $N^{3U}H$), 8.25 (s, 1H, H^{6U}), 7.63 (d, J = 7.9, 2H, H^f, H^g), 7.52 (d, J = 7.9, 2H, H^d, H^e), 6.92 (bs, 2H, $C^{6A}NH_2$), 6.01 (bs, 2H, $C^{2A}NH_2$), 4.09 (t, J = 6.9 Hz, 2H, $N^{9A}CH_2C_9H_{19}$), 3.71 (t, J = 6.9 Hz, 2H, $N^{1U}CH_2C_9H_{19}$), 1.97-1.87 (m, 2H, $N^{8A}CH_2CH_2C_8H_{17}$), 1.84-1.72 (m, 2H, $N^{1U}CH_2CH_2C_8H_{17}$), 1.34-1.11 (m, 28H, $N^{9A}C_2H_4C_7H_{14}CH_3$, $N^{1U}C_2H_4C_7H_{14}CH_3$), 0.87-0.79 (m, 6H, $N^{9A}C_9H_{18}CH_3$, $N^{1U}C_9H_{18}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.8, 161.2, 156.1, 152.0, 151.6, 149.9, 149.5, 131.6, 131.3, 129.2, 123.4, 120.6, 96.9, 92.1, 91.3, 85.3, 81.8, 48.1, 42.3, 31.23, 31.22, 28.95, 28.93, 28.88, 28.85, 28.82, 28.6, 28.5, 28.4, 28.3, 25.9, 25.7, 22.04, 22.01, 13.9.

HRMS (ESI+): Calculated for $C_{39}H_{53}N_8O_2$ $[M+H]^+$: 665.4285. Found: 665.4268.

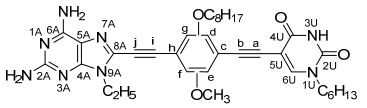


AU2. **AU2** was prepared following *Standard Procedure F* with $\mathbf{AU1_{Alk10}}$ (8.2 mg, 0.03 mmol), the mixture of isomers $\mathbf{U1_{AU2}}$ (50.0 mg, 0.08 mmol), CuI (0.05 mg, 0.0002 mmol) and $Pd(PPh_3)_4$ (0.6 mg, 0.0005 mmol) in the THF/NEt_3 mixture (2 mL). The reaction was completed overnight. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. The mixture of two regioisomers **AU2** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording the mixture **AU2** as a yellow solid (12.8 mg, 59%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 12.51 (bs, 1H, $N^{3U}H$), 12.23 (bs, 1H, $N^{3U}H$), 7.62-7.52 (m, 2H, H^f), 7.29 (s, 2H, H^{6U}), 7.11-7.03 (m, 2H, H^d), 6.47 (bs, 4H, $C^{6A}NH_2$), 5.65 (bs, 4H, $C^{2A}NH_2$), 4.27-4.15 (m, 4H, $N^{9A}CH_2C_9H_{19}$), 4.07-3.94 (m, 4H, $OCH_2C_7H_{15}$), 3.90-3.85 (m, 6H, OCH_3), 3.83-3.72 (m, 4H, $N^{1U}CH_2C_9H_{19}$), 1.97-1.66 (m, 8H, $N^{9A}CH_2CH_2C_8H_{17}$, $-OCH_2CH_2C_6H_{13}$), 1.48 (m, 4H, $N^{1U}CH_2CH_2C_8H_{17}$), 1.43-1.14 (m, 38H, $N^{9A}C_2H_2C_7H_{14}CH_3$, $N^{1U}C_2H_2C_7H_{14}CH_3$, $OC_2H_4C_5H_{10}CH_3$), 0.96-0.78 (m, 9H, $N^{9A}C_9H_{18}CH_3$, $N^{1U}C_9H_{18}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.8, 161.7, 161.1, 156.0, 156.0, 153.9, 153.2, 153.1, 152.9, 151.6, 149.9, 149.2, 148.8, 129.7, 116.4, 115.2, 114.6, 114.0, 113.9, 113.8, 111.1, 110.8, 97.5, 97.4, 88.22, 88.18, 79.2, 69.2, 68.8, 56.3, 56.2, 48.1, 48.0, 42.3, 31.28, 31.25, 31.23, 29.0, 28.91, 28.89, 28.78, 28.75, 28.73, 28.68, 28.64, 28.62, 28.59, 28.56, 28.48, 28.39, 28.37, 26.0, 25.7, 25.5, 25.4, 22.11, 22.09, 22.08, 22.06, 13.93, 13.90, 13.86.

HRMS (ESI+): Calculated for $C_{48}H_{70}N_8O_4$ $[M+H]^+$: 822.5515. Found: 822.5515.



AU3. **AU3** was prepared following *Standard Procedure F*. $\mathbf{AU1_{Alk10}}$ (18.2 mg, 0.09 mmol), the mixture of isomers $\mathbf{U1_{AU3}}$ (157 mg, 0.27 mmol), CuI (0.2 mg, 0.001 mmol) and $Pd(PPh_3)_4$ (2.1 mg, 0.002 mmol) were suspended in the THF/NEt_3 mixture (4 mL). The reaction was completed overnight. Then, the reaction mixture was filtered through a celite plug, washed with THF and concentrated under reduced pressure. The mixture of two regioisomers **AU3** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording the mixture **AU3** as a yellow solid (44.9 mg, 76%).

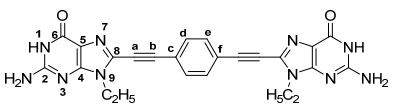
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 11.61 (bs, 2H, $N^{3U}H$), 8.20 (s, 1H, H^{6U}), 8.14 (s, 1H, H^{6U}), 7.21 (s, 1H, H^f), 7.18 (s, 1H, H^f), 7.09 (s, 1H, H^d), 7.08 (s, 1H, H^d), 6.87 (bs, 4H, $C^{6A}NH_2$), 6.00 (bs, 4H, $C^{2A}NH_2$), 4.14 (q, *J* = 7.0 Hz, 4H, $N^{9A}CH_2C_9H_{19}$), 4.05 (q, *J* = 7.0 Hz, 4H, $OCH_2C_7H_{15}$), 3.86 (s, 3H, OCH_3), 3.84 (s, 3H, OCH_3), 3.72 (t, *J* = 7.0 Hz, 4H, $N^{1U}CH_2C_9H_{19}$), 1.77-1.67 (m, 4H, $OCH_2CH_2C_6H_{13}$), 1.66-1.54 (m, 4H, $OC_2H_4CH_2C_5H_{11}$), 1.51-1.40 (m, 4H, $N^{1U}CH_2CH_2C_4H_9$), 1.36 (t, *J* = 7.0 Hz, 6H, $N^{9A}CH_2CH_3$), 1.32-1.17 (m, 28H, $N^{1U}C_2H_2C_3H_6CH_3$, $OC_3H_6C_4H_8CH_3$), 0.89-0.79 (m, 12H, $N^{1U}C_5H_{10}CH_3$, $OC_7H_{14}CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.8, 161.7, 161.2, 156.09, 156.07, 154.0, 153.2, 152.9, 151.4, 151.3, 149.9, 149.2, 148.9, 129.20, 129.19, 116.4, 115.2, 115.0, 114.6, 114.00, 113.95, 113.8, 111.1, 110.8, 97.5, 97.4, 89.5, 88.3, 88.2, 84.9, 79.2, 69.3, 68.8, 56.32, 56.27, 56.0, 48.08, 48.06, 37.3, 31.3, 31.2, 30.83, 30.81, 30.7, 28.75, 28.72, 28.69, 28.65, 28.39, 28.37, 25.50, 25.49, 25.41, 22.11, 22.06, 21.95, 18.6, 15.0, 14.9, 13.92, 13.88, 13.86.

HRMS (ESI+): Calculated for $C_{36}H_{47}N_8O_4$ $[M+H]^+$: 655.3714. Found: 655.3730.

Symmetric monomers.

GG monomers.



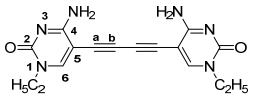
GG1. **GG1** was prepared following the *Standard Procedure F* with **G1$_{Alk2}$** (41 mg, 0.20 mmol), diiodobenzene **B1** (30 mg, 0.09 mmol), CuI (0.2 mg, 0.001 mmol) and $Pd(PPh_3)_4$ (2.1 mg, 0.002 mmol) in the DMF/NEt_3 mixture (5 mL). The reaction was completed overnight. Then, the resulting insoluble solid in suspension was filtered and washed with THF, acetonitrile and methanol, affording **GG1** as an ochre solid (37.6 mg, 86%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.71 (bs, 2H, N^1H), 7.71 (s, 4H, H^d, H^e), 6.62 (bs, 4H, C^2NH_2), 4.08 (q, *J* = 7.2 Hz, 4H, NCH_2CH_3), 1.35 (m, 6H, NCH_2CH_3).

HRMS (MALDI, dithranol): Calculated for $C_{24}H_{21}N_{10}O_2$ $[M+H]^+$: 481.1843. Found: 481.1859.

^{13}C NMR was not recorded due to the insolubility of **GG1** in any of the deuterated solvents available.

CC monomers.

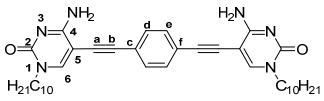


CC1. **CC1** was obtained as a byproduct after subjecting **C1$_{Alk2}$** to Sonogashira coupling conditions with **C1$_{Alk2}$** (50 mg, 0.31 mmol), CuI (1.2 mg, 0.006 mmol) and $Pd(PPh_3)_4$ (7.1 mg, 0.006 mmol) in a THF/NEt_3 mixture (5 mL). **CC1** precipitated in the mixture and was washed with cyclohexane and acetonitrile. **CC1** was obtained as a dark grey solid (39 mg, 39%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 9.47 (bs, 2H, C^2NH_2), 9.35 (bs, 2H, C^2NH_2), 8.68 (s, 2H, H^6), 3.33 (q, *J* = 7.2 Hz, 4H, $N^1CH_2CH_3$), 1.22 (t, *J* = 7.2 Hz, 6H, $N^1CH_2CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 159.7, 155.2, 146.8, 120.0, 80.1, 73.6, 45.2, 13.7.

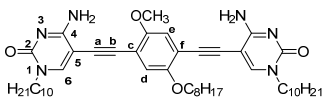
HRMS (ESI+): Calculated for $C_{16}H_{17}N_6O_2$ $[M+H]^+$: 325.1407. Found: 325.1420.



CC2. **CC2** was obtained as a byproduct in the preparation of product **C1$_{GC1}$**. **CC2** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording **CC2** as an ochre solid (178 mg, 11%).

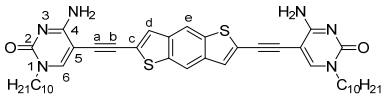
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.15 (s, 2H, H^6), 7.66 (bs, 2H, C^2NH_2), 7.66 (s, 4H, H^d, H^e), 6.99 (bs, 2H, C^2NH_2), 4.32 (t, *J* = 7.2 Hz, 2H, $N^1CH_2C_9H_{19}$), 3.69 (t, *J* = 7.0 Hz, 4H, $N^1CH_2C_9H_{19}$), 1.59 (m, 4H, $N^1CH_2CH_2C_8H_{17}$), 1.24 (m, 28H, $N^1C_2H_4C_7H_{14}CH_3$), 0.84 (t, *J* = 7.0 Hz, 6H, $N^1C_9H_{18}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 159.5, 150.6, 146.1, 131.8, 122.0, 97.5, 91.3, 51.3, 31.8, 29.7, 29.4, 29.3, 29.2, 29.0, 28.8, 26.2, 22.6, 14.00.
HRMS (ESI+): Calculated for $C_{38}H_{53}N_6O_2$ $[M+H]^+$: 625.4224. Found: 625.4285.



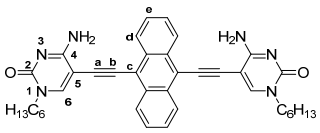
CC3. **CC3** was obtained as a byproduct in the preparation of product **C1$_{GC2}$**. **CC3** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording **CC3** as a pale solid (5 mg).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.80 (bs, 1H, C^{4}NH-*H*), 7.53 (s, 2H, *H*^{6}), 7.03 (bs, 1H, C^{4}NH-*H*), 6.90 (s, 2H, *H*^{d}, *H*^{e}), 6.30 (bs, 1H, C^{4}NH-*H*), 6.17 (bs, 1H, C^{4}NH-*H*), 4.02 (m, 2H, OC*H*$_2$C$_7$H$_{15}$), 3.93 (s, 3H, OC*H*$_3$), 3.79 (m, 4H, N^{1}C*H*$_2$C$_9$H$_{19}$), 1.88-1.66 (m, 6H, OCH$_2$C*H*$_2$C$_6$H$_{13}$, N^{1}CH$_2$C*H*$_2$C$_8$H$_{17}$), 1.46 (m, 2H, OC$_2$H$_4$C*H*$_2$C$_5$H$_{11}$), 1.41-1.18 (m, 36H, N^{1}C$_2$H$_2$C$_7$*H*$_{14}$CH$_3$, OC$_3$H$_6$C$_4$*H*$_8$CH$_3$), 0.87 (m, 9H, N^{1}C$_9$H$_{18}$C*H*$_3$, OC$_7$H$_{14}$C*H*$_3$).
^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.8, 164.6, 154.9, 153.7, 153.3, 146.9, 114.5, 113.4, 112.8, 112.5, 92.82, 92.75, 90.73, 90.69, 86.8, 86.5, 69.4, 56.4, 50.8, 50.7, 31.84, 31.78, 29.49, 29.45, 29.32, 29.28, 29.24, 26.6, 26.0, 22.6, 14.1.
HRMS (ESI+): Calculated for $C_{47}H_{71}N_6O_4$ $[M+H]^+$: 783.5531. Found: 783.5568.



CC4. **CC4** was obtained as a byproduct in the preparation of product **C1$_{GC8}$**. **CC4** precipitated in the medium and was washed first with THF and then with acetonitrile, affording **CC4** as a yellow solid (50 mg, 0.1%).

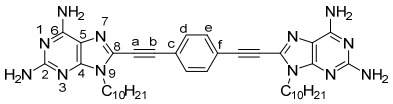
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.47 (s, 2H, *H*^{d}), 8.31 (s, 2H, *H*^{e}), 7.80 (m, 2H, C^{2}NH-*H*, *H*^{6}), 7.42 (bs, 1H, C^{2}NH-*H*), 3.71 (t, *J* = 7.0 Hz, 4H, N^{1}C*H*$_2$C$_9$H$_{19}$), 1.60 (m, 4H, N^{1}CH$_2$C*H*$_2$C$_8$H$_{17}$), 1.29-1.21 (m, 28H, N^{1}C$_2$H$_4$C$_7$*H*$_{14}$CH$_3$), 0.84 (m, 6H, N^{1}C$_9$H$_{18}$C*H*$_3$).
MS (ESI+): 737.38 $[M+H]^+$.



CC5. **CC5** was obtained as a byproduct in the preparation of product **C1$_{GC9}$**. **CC5** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (30:1), affording **CC5** as a red solid (10 mg, 2%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 8.58-8.51 (m, 4H, *H*^{d}, *H*^{e}), 7.75 (s, 2H, *H*^{6}), 7.69-7.65 (m, 4H, *H*^{f}, *H*^{g}), 6.83 (bs, 2H, C^{4}N*H*$_2$), 5.95 (bs, 2H, C^{4}N*H*$_2$), 3.86 (t, *J* = 7.0 Hz, 2H, N^{1}C*H*$_2$C$_5$H$_{11}$), 1.9-1.7 (m, 4H, N^{1}CH$_2$C*H*$_2$C$_4$H$_9$), 1.5-1.2 (m, 16H, N^{1}C$_2$H$_4$C$_4$*H*$_8$CH$_3$), 1.0-0.8 (m, 6H, N^{1}C$_5$H$_{11}$C*H*$_3$).
MS (ESI+): 613.3 $[M+H]^+$.

AA monomers.

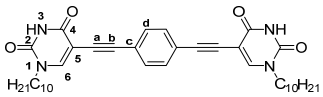


AA1. **AA1** was prepared following *Standard Procedure F* with **A1$_{Alk10}$** (0.42 g, 1.33 mmol), diiodobenzene **B1** (0.20 g, 0.61 mmol), CuI (11.6 mg, 0.006 mmol) and $Pd(PPh_3)_4$ (14.0 mg, 0.012 mmol) suspended in the THF/NEt_3 mixture (30 mL). The reaction was completed overnight. Then, the resulting insoluble solid in suspension was filtered and washed with THF, acetonitrile and chloroform, affording **AA1** as a dark yellow solid (42.9 mg, 10%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 9.40 (bs, 2H, C^{6}N*H*$_2$), 8.66 (bs, 2H, C^{6}N*H*$_2$), 7.77 (s, 4H, -Ar), 7.63 (bs, 4H, C^{2}N*H*$_2$), 4.15 (m, 4H, NC*H*$_2$C$_9$H$_{19}$), 1.79 (m, 4H, NCH$_2$C*H*$_2$C$_8$H$_{17}$), 1.4-1.0 (m, 28H, NC$_2$H$_4$C$_7$*H*$_{14}$CH$_3$), 0.79 (m, 6H, NC$_9$H$_{19}$C*H*$_3$).
^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 152.8, 132.8, 132.2, 121.6, 112.0, 93.4, 80.7, 43.2, 31.3, 28.9, 28.7, 28.4, 25.8, 22.1, 13.9.

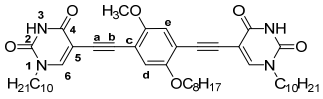
HRMS (ESI+): Calculated for $C_{40}H_{55}N_{12}$ $[M+H]^+$.: 703.4667. Found: 703.4679.

UU monomers.



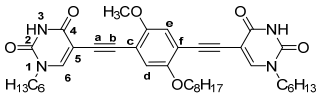
UU1. **UU1** was obtained as a byproduct in the preparation of product **U1$_{AU1}$**. **UU1** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (80:1), affording **UU1** as a white solid (4 mg).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.62 (bs, 2H, N^3*H*), 8.21 (s, 2H, *H*^{6}), 7.45 (s, 4H, *H*^{d}, *H*^{e}), 3.93 (m, 4H, N^{1}C*H*$_2$C$_9$H$_{19}$), 1.59 (m, 4H, N^{1C}CH$_2$C*H*$_2$C$_8$H$_{17}$), 1.38-1.18 (m, 28H, N^{1}C$_2$H$_4$C$_7$*H*$_{14}$CH$_3$), 0.83 (t, *J* = 7.0 Hz, 6H, N^{1}C$_9$H$_{18}$C*H*$_3$).
^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.3, 150.6, 148.0, 131.6, 122.4, 100.5, 95.1, 80.2, 50.4, 31.8, 29.4, 29.3, 29.2, 29.0, 28.9, 26.3, 22.6, 14.0.
HRMS (ESI+): Calculated for $C_{38}H_{51}N_4O_4$ $[M+H]^+$: 627.3904. Found: 627.3928.



UU2. **UU2** was obtained as a byproduct in the preparation of product **U1$_{AU2}$**. **UU2** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1), affording **UU2** as a pale solid (26 mg, 6%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 9.16-9.04 (bs, 2H, N^3*H*), 7.55 (s, 1H, *H*^{6}), 7.48 (s, 1H, *H*^{6}), 7.00 (s, 1H, *H*^{d}), 6.99 (s, 1H, *H*^{e}), 3.98 (t, *J* = 7.0 Hz, 2H, OC*H*$_2$C$_7$H$_{15}$), 3.85 (s, 3H, OC*H*$_3$), 3.75 (t, *J* = 7.0 Hz, 4H, NC*H*$_2$C$_9$H$_{19}$), 1.81 (m, 2H, OCH$_2$C*H*$_2$C$_6$H$_{13}$), 1.70 (m, 4H, NCH$_2$C*H*$_2$C$_8$H$_{17}$), 1.48 (m, 2H, OC$_2$H$_4$C*H*$_2$C$_5$H$_{11}$), 1.37-1.18 (m, 36H, NC$_2$H$_2$C$_7$*H*$_{14}$CH$_3$, OC$_3$H$_6$C$_4$*H*$_8$CH$_3$), 0.87 (t, *J* = 7.0 Hz, 9H, NC$_9$H$_{18}$C*H*$_3$, OC$_7$H$_{14}$C*H*$_3$).
^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 161.8, 161.6, 153.7, 153.4, 149.8, 146.8, 146.5, 117.3, 115.2, 113.7, 112.9, 100.2, 100.0, 90.3, 90.1, 85.6, 85.5, 69.6, 56.4, 49.3, 31.8, 29.4, 29.3, 29.2, 29.1, 26.4, 26.3, 25.9, 22.6, 14.0.
MS (ESI+): 785.5 $[M+H]^+$.



UU3. **UU3** was obtained as a byproduct in the preparation of product **U1$_{AU3}$**. **UU3** was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1), affording **UU3** as a pale solid (15 mg, 5%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 11.63 (bs, 2H, N^3*H*), 8.18 (s, 1H, *H*^{6}), 8.12 (s, 1H, *H*^{6}), 7.02 (s, 1H, *H*^{d}), 7.01 (s, 1H, *H*^{e}), 3.99 (t, *J* = 7.0 Hz, 2H, OC*H*$_2$C$_7$H$_{15}$), 3.79 (s, 3H, OC*H*$_3$), 3.71 (t, *J* = 7.0 Hz, 4H, NC*H*$_2$C$_5$H$_{11}$), 1.70 (m, 2H, OCH$_2$C*H*$_2$C$_6$H$_{13}$), 1.59 (m, 4H, NCH$_2$C*H*$_2$C$_4$H$_9$), 1.43 (m, 2H, OC$_2$H$_4$C*H*$_2$C$_5$H$_{11}$), 1.36-1.15 (m, 20H, NC$_2$H$_2$C$_3$*H*$_6$CH$_3$, OC$_3$H$_6$C$_4$*H*$_8$CH$_3$), 0.88-0.80 (t, *J* = 7.0 Hz, 9H, NC$_5$H$_{10}$C*H*$_3$, OC$_7$H$_{14}$C*H*$_3$).
^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 161.9, 161.8, 153.3, 152.9, 149.9, 149.0, 148.7, 117.2, 115.0, 113.4, 112.6, 97.6, 97.5, 88.4, 88.3, 87.62, 87.57, 69.1, 56.1, 48.1, 31.3, 30.8, 28.8, 28.7, 28.4, 25.4, 22.1, 22.0, 14.0, 13.9.
MS (ESI+): 673.4 $[M+H]^+$.

Chapter 3.

Molecular Self-Assembly at the Solid-Vacuum Interface.

The STM experiments under UHV conditions described in this Chapter were performed at the *Instituto IMDEA Nanociencia* (Madrid) in collaboration with the group of Dr. Roberto Otero, Dr. José-María Gallego and Dr. David Écija. Theoretical calculations were carried out at the *Universidad Autónoma de Madrid* in collaboration with Prof. Fernando Martín and Dr. Yang Wang.

3.1. Introduction. Molecular imaging by STM at the solid-vacuum interface.

As it has been introduced in the previous Chapter 2, molecular imaging by STM can be achieved under different conditions and, in this third Chapter, the focus will be put on the study of the self-assembly process of our dinucleobase monomers under UHV. Despite studying similar molecules than the ones introduced in Chapter 2, the decision to include this part of the work in a different Chapter of this Thesis was taken motivated by the existence of four main differing characteristics.

(**1**) In Chapter 2, the self-assembly study was carried out at the solid-liquid interface and in ambient conditions, where the molecules enjoy from a large conformational freedom, dynamics play an important role and solvent molecules can even participate in the assembly. Here, our molecules will be studied at the solid-vacuum interface, under UHV (10^{-10} mbar), at ultra-clean conditions, and thus excluding the influence of the solvent.

(**2**) The second variation is the substrate. While in the second Chapter HOPG was used, different coinage metal substrates (Au (111) and Ag (111)) will be considered in this Chapter.

(**3**) A third main difference lies in the deposition technique. Previously, a solution of the monomer was drop-casted onto freshly cleaved HOPG, whereas now the molecular species will be sublimated and deposited on the surface by organic molecular beam epitaxy.

(**4**) Finally, not all the species adapt well to UHV conditions, since the technique implies sublimation of the sample, and thus potential thermal cracking of the species needs to be considered. Therefore, molecules with lower molecular weight will be targeted in Chapter 3.

The basic principle and operation of a scanning tunneling microscope, explained in the previous Chapter, is also applicable here. The UHV-system consists of a preparation chamber, housing all the preparation and deposition facilities, and a measurement chamber, hosting a low temperature scanning tunneling microscope (Figure 84). Preparation and measurement chambers can be pumped independently and the resulting base pressure after bake-out is below 5 x 10^{-10} mbar. In addition, there is a load-lock that allows the easy introduction of samples and tips into the preparation chamber without breaking the vacuum. During the STM measurements, the pumps are frequently turned off

in order to reduce the mechanical noise. Various molecular evaporators, working with resistive heating of a tungsten filament, are attached to the preparation chamber.

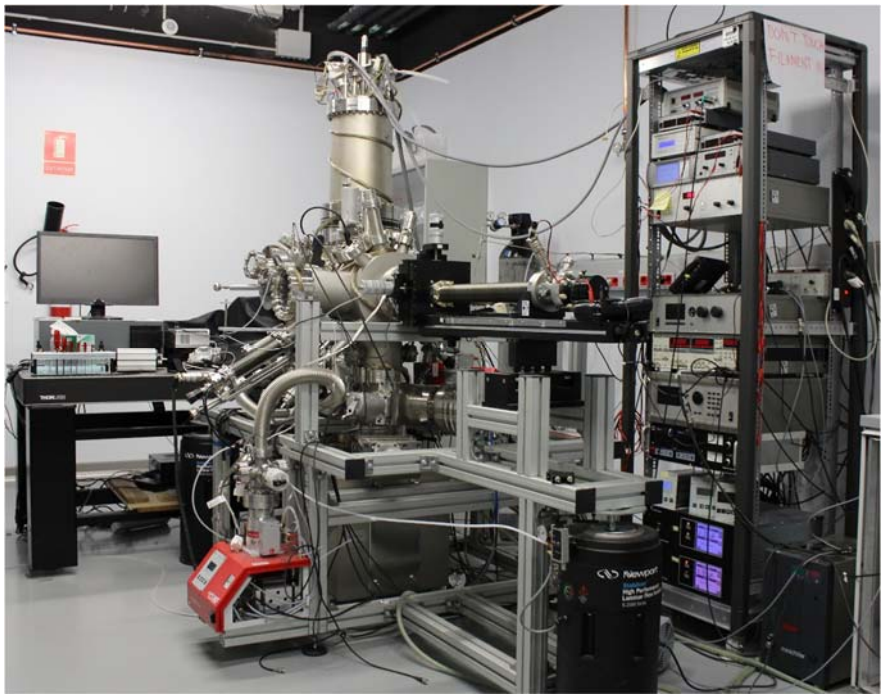


Figure 84. Photograph of the Omicron LT-STM setup used for the measurements detailed in this Chapter.

The 2D system can be described in terms of adsorption, mobility and lateral interactions. These parameters are a function of the atomic environment, the chemical nature and the symmetry of the substrate (usually noble metals and Cu). Typically, the reactivity toward adsorbed molecules decreases in the order: Cu (111) >> Ag (111) > Au (111); while the mobility increases.[230] When deposited on atomically clean Au, Ag or Cu substrates, a same molecule can form different patterns. When this occurs, it is usually directly related to the rotation angles of the molecule substituents, which depend on surface chemistry and orientation. The interaction with the surface atomic lattice affects the motion and conformation of the adsorbed molecules, whose characteristics can even be modified.[231] For instance, H-bonding energetics, distances, and geometrical alignment at metals do not necessarily reflect the parameters encountered in a 3D self-assembly medium (in solution or solid state environments). Additionally, the surface can present some catalytic activity and result in chemical modifications of specific functional groups.

[230] a) J. K. Gimzewski, C. Joachim, *Science* **1999**, *283*, 1683–1688; b) S. M. Barlow, R. Raval, *Surf. Sci. Rep.* **2003**, *50*, 201–341; c) J. V. Barth, J. Weckesser, C. Cai, P. Günter, L. Bürgi, *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 1230–1234; d) F. Rosei, M. Schunack, Y. Naitoh, P. Jiang, A. Gourdon, *Prog. Surf. Sci.* **2003**, *71*, 95–146.
[231] R. Otero, J. M. Gallego, A. L. Vázquez de Parga, N. Martín, R. Miranda, *Adv. Mater.* **2011**, *23*, 5148–5176.

In a metallic substrate prepared for STM measurements, the atomic coordination and the electronic structure in the surface region is changed relative to the bulk. In order to minimize the surface free energy, a small relaxation (expansion or contraction) in the equilibrium positions of the surface atoms takes place for the majority of clean metals. In the particular case of Au, the atomic positions are drastically rearranged in the surface region: it is said to be "reconstructed". It is the result of a balance between the energy gain, associated with the increased atomic density in the surface plane; and the energy loss upon reconstructing, due to the disruption or stretching of the bonds between the mismatched top and sub-surface layers. Barth *et al.*[232] showed how gold reconstruction resulted in periodic transitions between two stacking domain structures that created 0.2 Å corrugation lines. The resulting zigzag superstructure affords a herringbone or chevron pattern (V-shaped) that changes orientation over distances of up to 250 Å. Thermal annealing of gold thin films affords a smoothened surface, with larger Au (111) crystalline terraces (∼300 nm) and a lower amount of Au clusters.

To deposit molecules under UHV conditions, organic molecular beam epitaxy is used, allowing highly pure films to grow as a crystalline overlayer.[233] Successful STM imaging requires balanced adsorbate-substrate interactions. On the one hand, they have to be strong enough to immobilize and visualize individual isolated molecules: too weak adsorbate-substrate interactions lead to a too high mobility and STM imaging becomes impossible. To overcome the problem of too high molecular mobility on a given substrate, one can control the temperature or increase the monolayer coverage. Diffusion is, nonetheless, also observed for higher monolayer coverages. On the other hand, when a too strong interaction immobilizes the molecules, self-assembly into ordered 2D layers can be prevented and only kinetically frozen situations are imaged. At the solid-gas interface, temperatures and partial pressures are key factors in the adsorption of the molecules on the surface. For single molecule spectroscopy, for instance, UHV and very low temperatures (realized by liquid helium cooling) are necessary, whereas for studies of extended periodic molecular network structures room temperature conditions might be sufficient. Low temperatures are useful to prevent undesired diffusion events and to enhance the stability of the tunneling junction. However, diffusion must be sufficient to form the self-assembled molecular networks. Annealing, that is increasing the temperature, is often necessary so that the deposited molecules become ordered in specific patterns and reach the thermodynamically stable state. The activation energy for diffusion is the barrier an adsorbate has to overcome on the potential energy surface to reach at least the neighboring adsorption site. Substituents usually increase the distance between the aromatic core and the substrate, resulting in a decrease in adsorbate-substrate interactions. Large molecules occupy a larger area on the substrate and they might find it more difficult to find a low-energy site. By varying the nature of the substrate where the

[232] J. V. Barth, H. Brune, G. Ertl, R. J. Behm, *Phys. Rev. B* **1990**, *42*, 9307–9318.
[233] S. R. Forrest, *Chem. Rev.* **1997**, *97*, 1793–1896.

molecules are deposited, one can also control diffusion phenomena. In the case of gold, the corrugation of the potential energy landscape for the molecules adsorbed on Au (111) is small enough to grant some molecular diffusion, while keeping the molecules adsorbed. Therefore, this substrate gives the molecules the possibility to form strong directional intermolecular H-bonds, for example. The corrugation of the potential energy for the adsorbed species is small compared to the energy gain from intermolecular interactions.

Under UHV, both individual atoms as well as larger molecules can be studied with higher spatial resolution, under well-defined conditions, on a large variety of substrates, over a wide temperature range and with control of the surface coverage. A proper control of molecular organization on surfaces, especially metallic substrates, requires a good understanding of molecule-surface interactions. In order to achieve such a challenging task, it is often mandatory to combine experiments with computational modeling. STM can frequently clarify basic structural aspects, but fails sometimes to unequivocally assign the corresponding correct structure. Molecular building blocks usually adsorb at specific sites because the functional groups sense the substrate atomic arrangement.[234] Density-functional theory (DFT) and other model calculations provide insight into the bonding characteristics.[235] In 2D H-bonded assemblies, the lengths of the non-covalent unions are frequently increased in comparison with 3D analogs, which in the simplest cases can be attributed to the preferential occupation of specific adsorption sites.[236] Moreover, length variations of the same 2D H-bond can be appreciated by changing the substrate.[237] Even if the corrugation of the potential energy for the monomers adsorbed on Au (111) is small compared to the energy gain from intermolecular H-bonding, the substrate influence may still be important.[238] Heteroatoms, such as oxygen, nitrogen or sulfur, can create coordinative covalent bonds to the surface through overlap of the lone pairs at the heteroatoms and the *d* states of the metal, without further disruption of any bond at the molecular backbone.[230b,234a,239] Theoretical models taking into account different

[234] a) A. Hauschild, K. Karki, B. C. C. Cowie, M. Rohlfing, F. S. Tautz, M. Sokolowski, *Phys. Rev. Lett.* **2005**, *94*, 036106; b) A. Nilsson, L. G. M. Pettersson, *Surf. Sci. Rep.* **2004**, *55*, 49–167; c) M. Eremtchenko, J. A. Schaefer, F. S. Tautz, *Nature* 2003, 425, 602–605.

[235] a) Y. Okuno, T. Yokoyama, S. Yokoyama, T. Kamikado, S. Mashiko, *J. Am. Chem. Soc.* **2002**, *124*, 7218–7225; b) M. Vladimirova, G. Trimarchi, A. Baldereschi, J. Weckesser, K. Kern, *Acta Mater.* **2004**, *52*, 1589–1596; c) M. Preuss, W. G. Schmidt, F. Bechstedt, *Phys. Rev. Lett.* **2005**, *94*, 236102; d) A. Alkauskas, A. Baratoff, C. Bruder, *Phys. Rev. B* **2006**, *73*, 165408.

[236] a) G. E. Thayer, J. T. Sadowski, F. M. Heringdorf, T. Sakurai, R. M. Tromp, *Phys. Rev. Lett.* **2005**, *95*, 256106; b) J. V. Barth, J. Weckesser, G. Trimarchi, M. Vladimirova, A. De Vita, C. Cai, H. Brune, P. Günter, K. Kern, *J. Am. Chem. Soc.* **2002**, *124*, 7991–8000.

[237] S. Clair, S. Pons, A. P. Seitsonen, H. Brune, K. Kern, J. V. Barth, *J. Phys. Chem. B* **2004**, *108*, 19392–19397; (b) L. M. A Perdigao, N. R. Champness, P. H. Beton, *Chem. Commun.* **2006**, *2006*, 538–540.

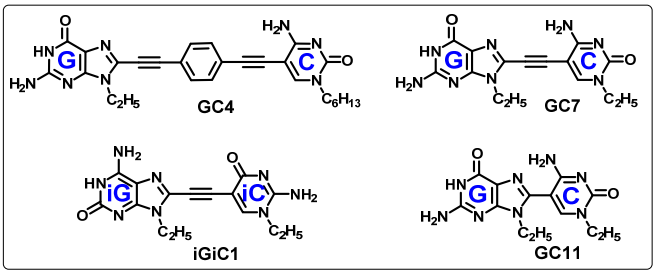
[238] a) R. Otero, M. Schöck, L. M. Molina, E. Laegsgaard, I. Stensgaard, B. Hammer, F. Besenbacher, *Angew. Chem. Int. Ed.* **2005**, *44*, 2270–2275; b) J. V. Barth, J. Weckesser, C. Cai, P. Günter, L. Bürgi, O. Jeandupeux, K. Kern, *Angew. Chem. Int. Ed.* **2000**, *39*, 1230–1234; c) J. A. Theobald, N. S. Oxtoby, M. A. Phillips, N. R. Champness, P. H. Beton, *Nature* **2003**, *424*, 1029–1031.

[239] a) T. Tseng, C. Urban, Y. Wang, R. Otero, S. L. Tait, M. Alcamí, D. Écija, M. Trelka, J. M. Gallego, N. Lin, M. Konuma, U. Starke, A. Nefedov, A. Langner, C. Wöll, M. Á. Herranz, F. Martín, N. Martín, K. Kern, R. Miranda, *Nature Chem.* **2010**, *2*, 374–379; b) K. Kwon, K. L. Wong, G. Pawin, L. Bartels, S. Stolbov, T. S. Rahman, *Phys. Rev. Lett.* **2005**, *95*, 166101; c) D.

parameters, as well as the interaction with the substrate, may be helpful for a clearer understanding of the self-assembly process in two dimensions.

Our contribution:

In this Chapter, we aim to investigate how lighter monomers, based on our linear design, may steer the formation of nanoporous cyclic tetramer networks on metallic surfaces under UHV conditions. The target molecules of this Chapter 3, all of them carrying self-assembling complementary bases in the edges, are shown in Scheme 21.



Scheme 21. Considered target monomers in Chapter 3.

Monomer **GC4** was introduced in the previous Chapter. This molecule resembles the structure of **GC3**, a monomer that we succeeded to self-assemble into large and stable porous networks of cyclic tetramers at the HOPG-liquid interface (see Chapter 2, section **2.2.1.**). Therefore, its study represents the most interesting one for us in order to compare molecular organization under both situations. Despite being heavier than the rest of the targets (MW = 496.6 g/mol), monomer **GC4** could, in principle, form stable cyclic tetramers on the metallic surface if sublimated.

We decided to start the investigation under UHV with **GC7**, a lighter monomer (MW = 340.4 g/mol) that should be more easily sublimated and whose supramolecular organization at the HOPG-liquid interface had been studied as well. Monomer **GC7** is again based on the GC Watson-Crick pair, both nucleobase derivatives are endowed with a short ethyl chain and separated by an ethynyl group. As seen in the previous Chapter, **GC7** was able to form macrocyclic networks onto HOPG at the solid-liquid interface. However, we concluded that the co-assembly with solvent molecules hampered the formation of dense porous domains and mainly row organizations were imaged (see Figure 67 in Chapter 2). At this point, in the absence of solvent molecules, **GC7** seemed an attractive first candidate to explore the self-assembly process on metallic substrates under UHV.

Wegner, R. Yamachika, Y. Wang, V. W. Brar, B. M. Bartlett, J. R. Long, M. F. Crommie, *Nano Lett.* **2008**, *8*, 131–135; d) M. Ortega-Lorenzo, C. J. Baddeley, C. Muryn, R. Raval, *Nature* **2000**, *404*, 376–379.

3.2. Results and Discussion.

3.2.1. Molecule GC7.

Prior to sublimation attempts, thermogravimetric analyses (TGA) were carried out under N_2 atmosphere in order to check **GC7** thermal stability. No weight loss was observed before reaching 250 ^{o}C (Figure 85). Above that temperature, the molecules began to thermally decompose.

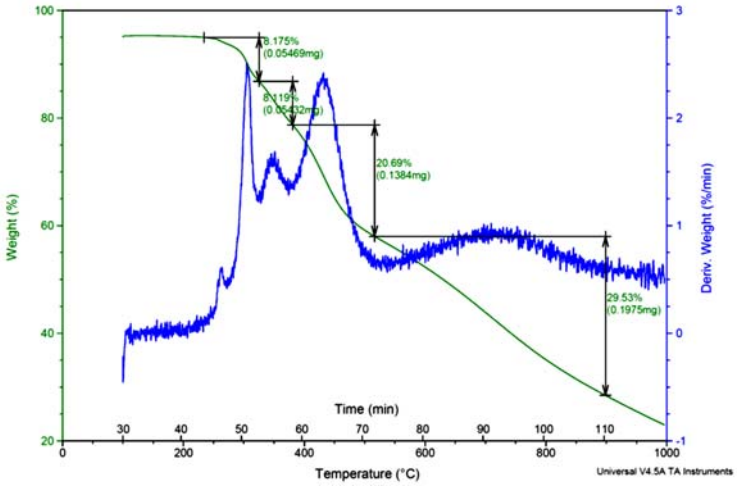


Figure 85. TGA of **GC7** under N_2 atmosphere.

Ag (111) and Au (111) were considered as metallic surfaces for the study of this molecule. The more reactive Ag (111) substrate did not afford a good coverage and isolated bright dots were observed (Figure 86).

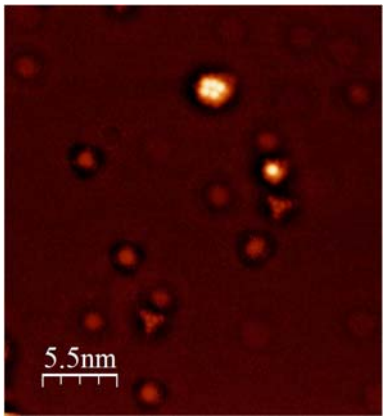


Figure 86. STM image of **GC7** on Ag (111). Tunneling parameters: V_t = 1 V; I_t = 100 pA.

Au (111) was instead used as the substrate. This metal surface is less reactive in general terms and it does not restrict as much the mobility of the molecules. Also, H-bonding is typically relatively unaffected by the interaction with the Au (111) surface. Below 200 ºC, no molecular deposition on the substrate was observed. Once this 200 ºC temperature was reached and after 5 min, triangular islands and stripes, composed of bright dots and rods that elongated across the gold terraces could be observed at low coverages (Figure 87). Thermal activation can trigger the formation of regular superstructures by rearranging the adsorbed molecular tectons.[240]

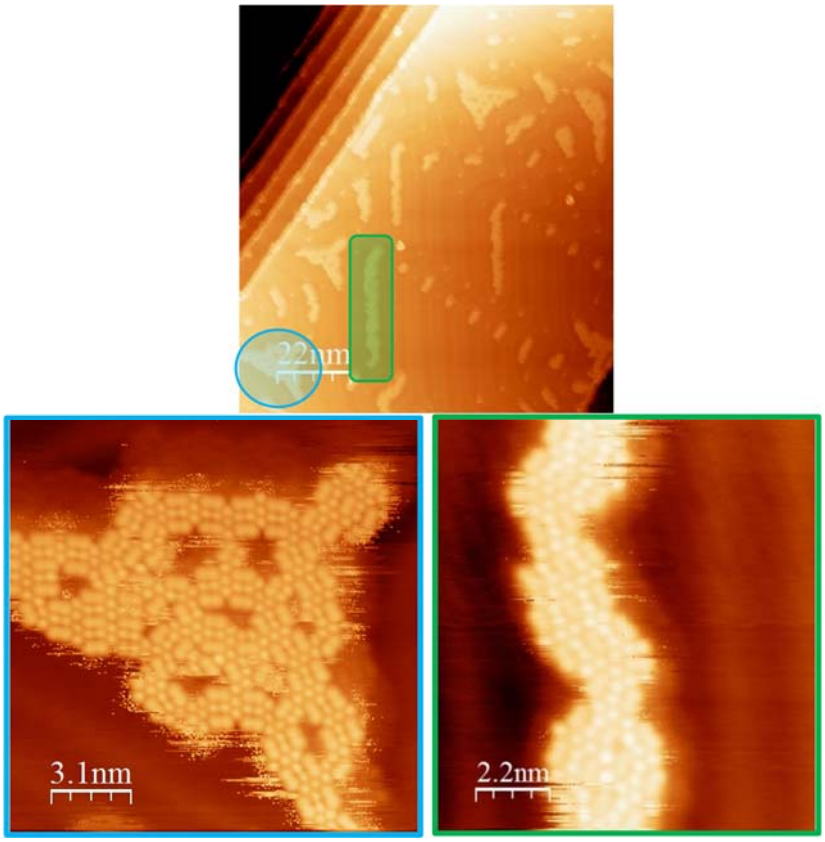


Figure 87. Low coverage STM image of **GC7** on Au (111). Tunneling parameters: V_t = -1.5 V; I_t = 15 pA.

By increasing the deposition, high coverage images were obtained. Well-resolved large domains of a regular porous network with C_3 symmetry were registered (Figure 88). One could see that the edges were favorable low energy adsorption sites. In these images, triangular-shaped pores (1), separated by 2.0 ±0.2 nm, could be seen (Figure 88b,c). Small dots (2 in Figure 88b) with a moderate brightness, bigger and brighter dots (3), as well as equally shiny rods (4) could be identified in the image. Each small dot (2) was approximately

[240] A. Kühnle, L. M. Molina, T. R. Linderoth, B. Hammer, F. Besenbacher, *Phys. Rev. Lett.* **2004**, *93*, 086101.

0.4 ±0.2 nm apart from its neighbour larger dot (3). This arrangement was repeated three times, following a 60° angle rotation and surrounding a bigger shiny dot (5). Let's note that the motifs 2; 3 and 4 reflected a 3:1 distribution in relation with motifs 1 and 5. In some cases, the pores were filled with molecules of unknown nature. The symmetry reduction associated with the 2D confinement brought along a rich variety of chirality phenomena.[230b,241] As seen on Chapter 2, depending on which face of the monomer was adsorbed on the surface, two different surface chiralities could be identified (Figure 88). Since the pores had a windmill-like blade shape, a clockwise (CW) and a counter-clockwise (CCW) sense of rotation could be defined.

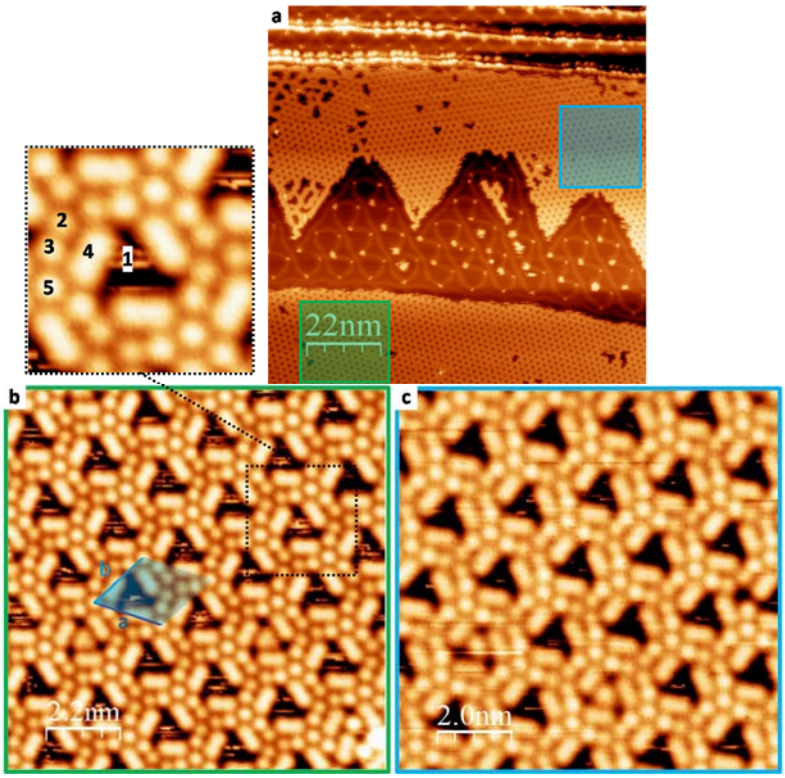


Figure 88. High coverage STM images of GC7 on Au (111) showing domains with different surface chirality. (**a**) Small scale image (tunneling parameters: V_t = -1.5 V; I_t = 5 pA) showing the coexistence of two domains with opposed surface chirality. (**b**) Large scale image corresponding to CCW pore image along with magnification (V_t = 0.3 V; I_t = 10 pA). The unit cell is marked in blue (a = 2.1 ±0.1 nm; b = 2.0 ±0.1 nm; α = 60 ±0.1°). (**c**) Large scale image corresponding to CW pore image, right (V_t = -0.5 V; I_t = 10 pA).

The self-assembled porous network was subjected to a voltage evolution experiment where the bias voltage was incremented from a negative value of -2.3 V to a positive voltage

[241] K. H. Ernst, *Top. Curr. Chem.* **2006**, *265*, 209–252.

of +2.5 V, while keeping the tunnel current constant. The different images recorded every 0.2 V increment, showed in a first moment, how the definition of the STM image was improved until V_t = 0.3 V. Then, the contrast in the images was lost and a thick bright network where only the triangular pores could be resolved was registered (Figure 89).

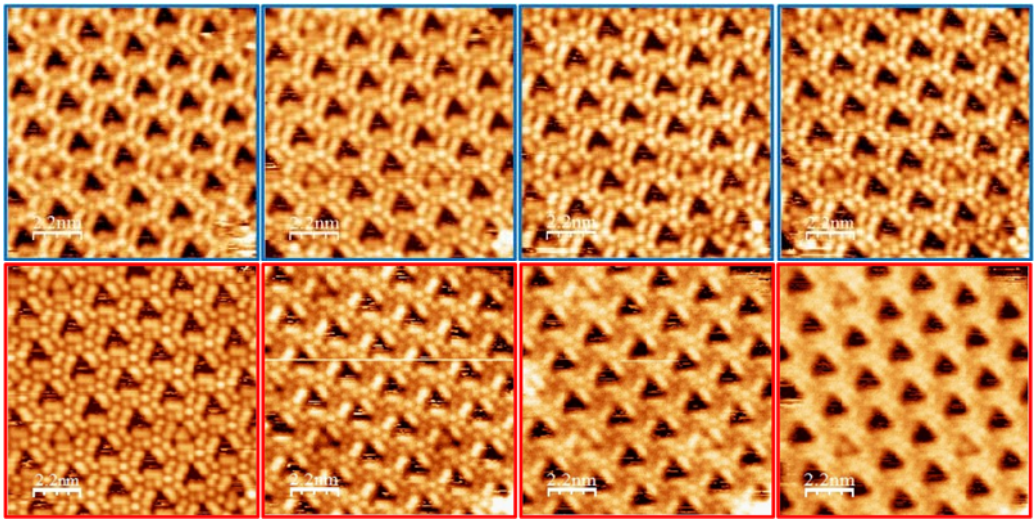


Figure 89. Voltage evolution experiment. Selected high coverage STM images of **GC7** on Au (111) (I_t = 10 pA) taken from V_t = -2.3 V to V_t = +1.8 V in +0.6 V steps. Blue squares indicate negative voltages and red, positive voltages.

Annealing experiments were carried out in order to see how the self-assembly responded to further thermal treatment. When increasing the temperature, the contour definition of the well-resolved shapes of the network got lost. One could still recognize the nucleation and growth process of this 2D crystal that started in triangular islands and elongated fiber-like structures, which we could not resolve (Figure 90).

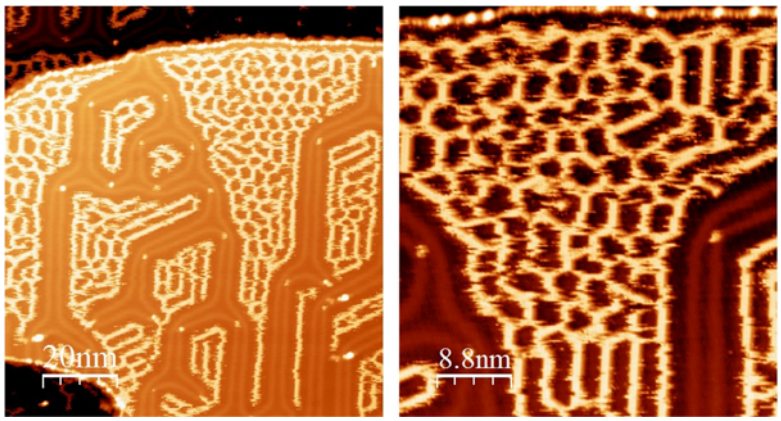


Figure 90. Annealing experiments. STM images of **GC7** on Au (111) after a short annealing process [T = 3.1 mV (~100 °C); t = 10 min]. Tunneling parameters: V_t = 2.5 V; I_t = 5 pA.

Although a stable 2D porous network was obtained, its features were totally different from those observed at the solid-liquid interface (Chapter 2, Figure 67). The symmetry and parameters did not fit the proposed model formed by tetrameric macrocycles, where G and C nucleobases formed a 90º angle through Watson-Crick molecular recognition and secondary H-bonding interactions between aminopyridine fragments further stabilized the self-assembled architecture (Figure 91). Instead, clear triangular patterns could be appreciated (Figure 88).

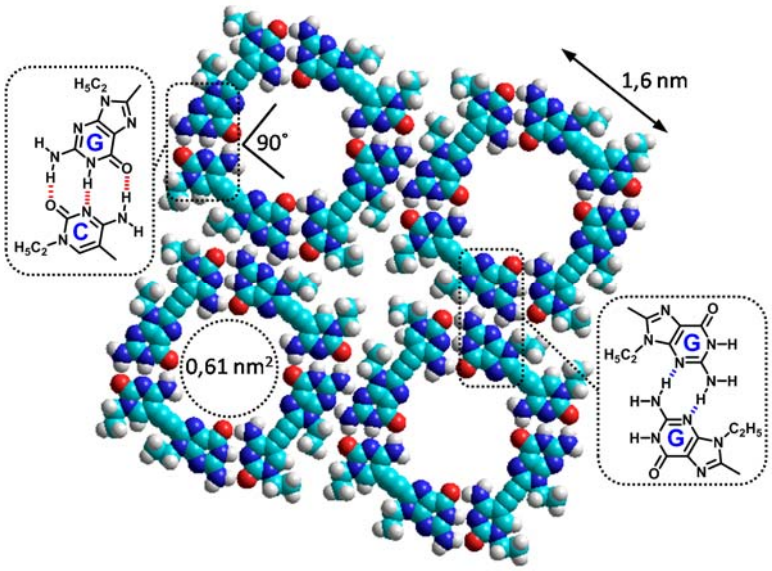


Figure 91. Anticipated model for **GC7** based on experimental data at the solid-liquid interface.

In order to shed light on the network obtained on Au (111) under UHV, we first examined the residue remaining in the crucible after deposition (Figure 92).

The ^{1}H NMR spectrum of this solid indicated the existence of a second product, not present before sublimation. The ratio between the starting compound and this novel product was calculated by signal integration as 72:28, respectively. The duplication of the signals corresponding to the alkyl chains, as well as the appearance of novel aromatic signals could indicate the event of a side-reaction taking place in the **GC7** solid during prolonged heating at 200 °C. The new species showed signals that could be attributed to the formation of a pyrrole ring by intramolecular nucleophilic attack of the C amino group to the adjacent triple bond, yielding **GpC** (Figure 92b).

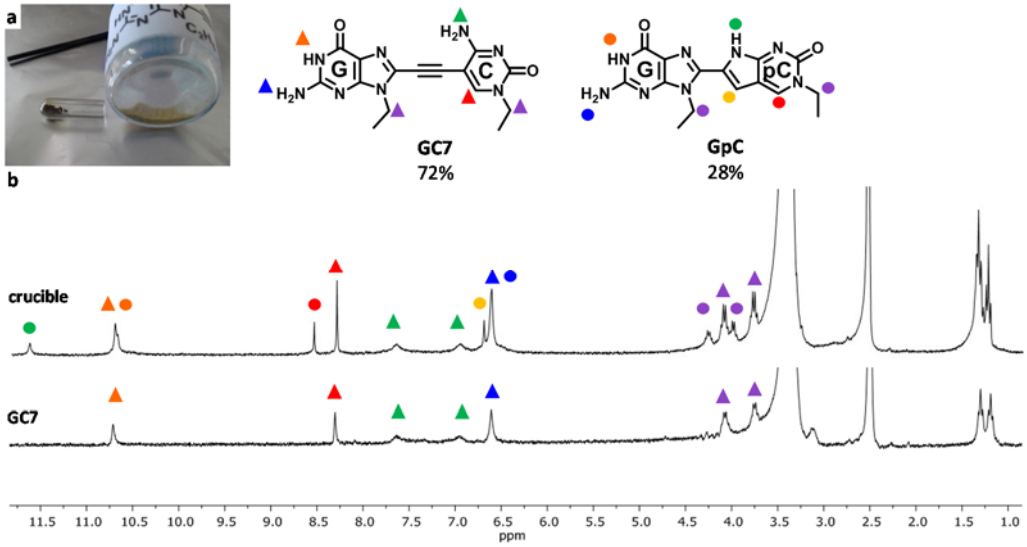
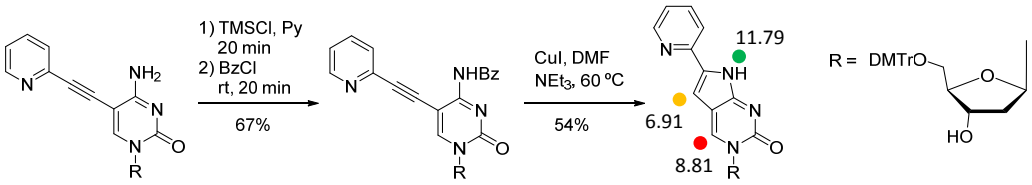


Figure 92. Comparative ^{1}H NMR spectra in DMSO-d_6 of the residue remaining in the crucible after deposition and pure GC7. (**a**) Photograph showing the starting material **GC7** (right); the crucible and the darker recovered solid (left). (**b**) ^{1}H NMR spectrum in DMSO-d_6 of the solid collected from the crucible, corresponding to a mixture containing 72% of **GC7** and a 28% of a byproduct (top). ^{1}H NMR spectrum in DMSO-d_6 corresponding to pure **GC7** (down).

Mei *et al.*[242] described how a similar starting product undergoes, in the presence of CuI, a cyclization reaction through the amino functionality, substituted with a suitable leaving group (Scheme 22). The electron lone pair of the nitrogen atom of the benzoyl-substituted amino group attacked the electrophilic carbon atom of the ethynyl group, forming a non-strained five-membered ring after elimination of the benzoyl moiety. Taking into account that our sample had been heated up to 200 °C for long periods of time; and the presence of traces of CuI from the final Sonogashira coupling reaction (see Experimental Part, section **2.4.2.**) cannot be excluded, **GC7** could have experienced this intramolecular cyclization reaction. Moreover, the chemical shifts described by Seela and coworkers[242] (Scheme 22) were similar to the ones found in the ^{1}H NMR spectrum of the crucible residue (Figure 92). In order to see if we could detect this thermal rearrangement, Differential Scanning Calorimetry (DSC) was carried out on **GC7**. Unfortunately, the DSC results did not provide any new information. Let's note that **GC7** and the cycled byproduct possess the same molecular weight, explaining no extraordinary change observation in TGA experiments around 78 °C (see Figure 85). It also explains why mass spectrometric analysis of the residue after sublimation revealed a main peak corresponding to the **GC7**/**GpC** molecular ion.

[242] H. Mei, I. Röhl, F. Seela, *J. Org. Chem.* **2013**, *78*, 9457–9463.



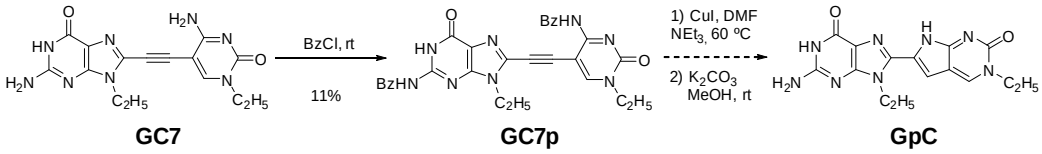
Scheme 22. Described intramolecular cyclization reaction along with the reported chemical shifts (in ppm) of the characteristic proton signals.[242]

At this point in which the STM images seemed to correspond to the adsorption on Au (111) of only one compound, but the residue collected from the crucible, after sublimation of the sample, contained a mixture of two products, it was not clear to us which molecule was deposited on the surface. This circumstance made thus the analysis of the STM images more difficult. In order to approach this issue, we considered two strategies that were undertaken in a parallel manner. The first one consisted in synthesizing the cycled byproduct **GpC**, with the aim of sublimating it and compare the STM images. In the second one, we addressed the chemical analysis of the sublimated residue.

Synthetic approach.

In order to corroborate the results exposed above, the synthesis of the cycled product **GpC** was attempted. In a first approach, **GC7** was progressively heated in DMF or deuterated DMSO and we followed the evolution of the process by ^{1}H NMR. At moderate temperatures only the starting material was observed, whereas at higher temperatures (> 120 ^{o}C) decomposition of the product took place, leading to a complex mixture that was difficult to analyze. It was then essayed to heat **GC7** at 60 ^{o}C in DMF together with CuI in the presence of NEt_3, mimicking the reaction conditions described by Mei.[242] Again, the starting product was recovered and an increase of the temperature of the reaction resulted in decomposition of the starting material.

In a second approach, the two steps of the cyclization reaction proposed by the group of Seela[242] were considered (Scheme 23). First the amino group was blocked with a benzoyl residue, resulting in the intermediate **GC7p**, which was isolated and characterized. Let's note that not only the C-amino group, but also the amino group of the G-derivative, reacted with benzoyl chloride. Therefore, after the intramolecular step in the presence of CuI and NEt_3 that resulted in the formation of the pyrrole system, the G-base benzoyl group was directly cleaved through the addition of potassium carbonate in methanol, after evaporation of the previous solvent (DMF). A complex mixture was obtained, which was difficult to purify because of the extraordinary insolubility of **GpC**, resulting in supplementary manipulation problems during reaction and purification protocols. Unfortunately, product **GpC** could not be satisfactorily isolated and identified in none of the separated fractions of the crude mixture.



Scheme 23. Attempted synthetic procedure toward product **GpC**.

Analysis of the residue.

Another strategy consisted in repeating the STM experiment, collecting the residue in the evaporator shutter after the sublimation protocol (18 h at 210 ºC) prior to deposition and analyze it. The low amount of collected sample made the analyses difficult, and unfortunately, no **GC7**, nor **GpC**, were identified by ^{1}H NMR and MS experiments in this complex mixture, which contained other non-identified species, having C and G fragments.

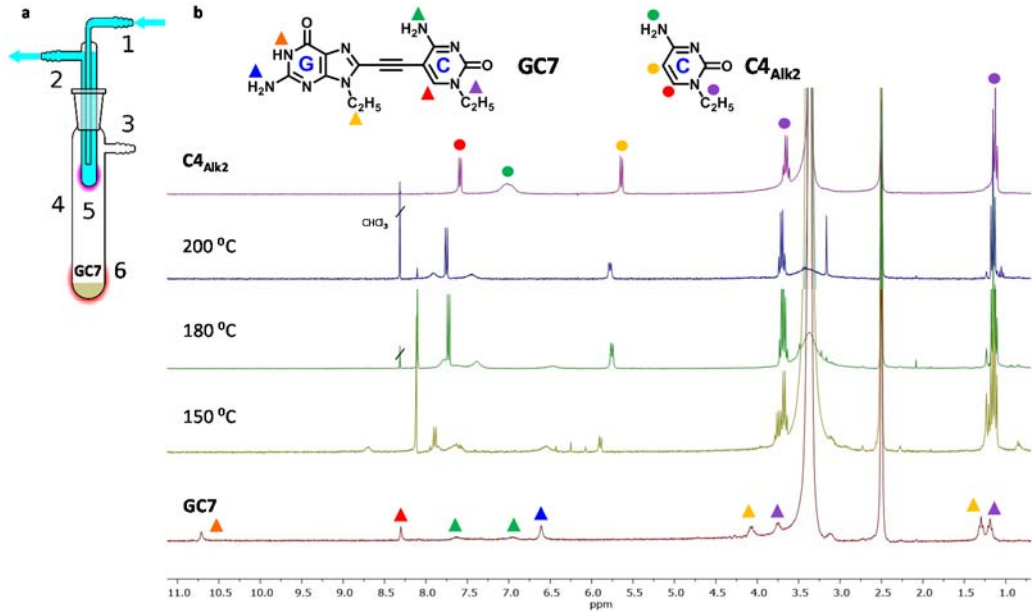


Figure 93. (**a**) Cold finger setup. The raw product (**GC7**) is in the bottom of the outer tube (4) which is heated (6) while under vacuum (through side-arm 3). The sublimated material collects (5) on the cold finger proper, cooled by a coolant (blue) circulated through ports 1 and 2.[243] (**b**) Comparative ^{1}H NMR spectra in DMSO-d_6 of the successive sublimated solids, recovered on the cold finger proper, after heating the sample 8 h at the indicated temperature.

In another attempt the sublimation conditions of the UHV technique for STM experiments were mimicked at our laboratories. A flask containing the ochre solid **GC7** was connected to a cold finger (Figure 93a), the setup was put under vacuum (10^{-1} mbar) and first heated at 100 ºC. No solid sublimated and the temperature was elevated to 150 ºC.

[243] www.wikipedia.org

After 8 hours, the sublimated collected material was analyzed by ^{1}H NMR and mainly corresponded to the starting material **GC7**. Two new doublets downfield and two other signals upfield, overlapping the C ethyl signals, could be distinguished. The remaining solid was again heated at 180 ^{o}C for 8 hours and the sublimated collected material was analyzed one more time by ^{1}H NMR. The spectrum was very similar to the previous one but the doublet signals, possibly corresponding to the cytosine residue **C4$_{Alk2}$**, had grown. Finally, the residual solid was heated at 200 ^{o}C for 8 hours and the collected material (darker this time) was analyzed by ^{1}H NMR (Figure 93b). The spectrum obtained seemed to mostly correspond to **C4$_{Alk2}$**. The ^{1}H NMR spectrum of the remaining solid corresponded to a mixture of **GC7** and **C4$_{Alk2}$**. Therefore, it seemed that prolongated heating of **GC7** at high temperatures resulted in partial thermal cracking. One of the products of this decomposition process was **C4$_{Alk2}$**, as also confirmed by MS spectrometry. These results did not match in any case what was previously encountered for the crucible residue.

As all our efforts to: (1) synthesize **GpC** and/or (2) analyze the sublimated residue, resulted unproductive, we tried to analyze the STM images, even if we were not sure about the nature of the adsorbed molecules on the gold surface. Let's underline that this uncertainty made from the model elucidation a great challenge.

It is interesting to note that the C *DAA* H-bonding pattern is not lost after the intramolecular cyclization reaction, so Watson-Crick-type pairing to G is still possible in **GpC**. However, the binding angle changes drastically from 90^o in **GC7** to 60^o in **GpC**, which would lead to trimeric, instead of tetrameric, cyclic assemblies (Figure 94). This species should be very stable as well, since it is also a small non-strained cycle, where numerous intermolecular stabilizing H-bonds intervene. Additionally, the **GpC** molecule has a single rotatable bond and a preorganized structure for ring closure.

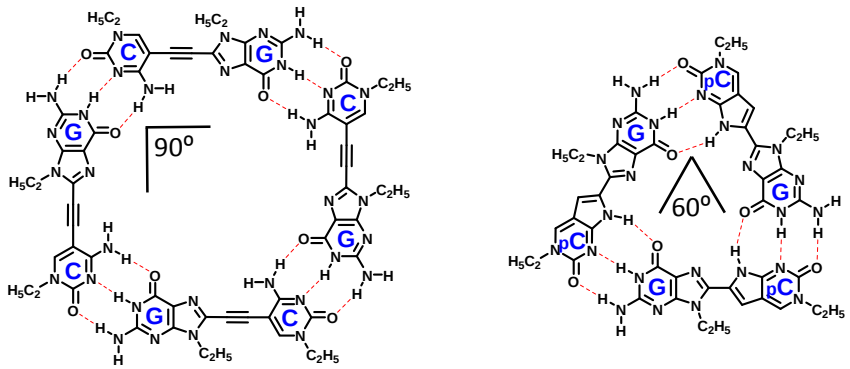


Figure 94. H-bonded tetrameric (left) and trimeric (right) cyclic assemblies formed by **GC7** and **GpC**, respectively.

Due to the dimensions and the C_3-symmetry of the observed patterns, it seemed reasonable to start with **GpC**. Therefore, the first proposed models considered that the STM images corresponded to the adsorption of this byproduct, where three molecules formed a H-bonded cyclic trimer (Figure 94, right). Preliminary DFT studies in the gas phase on **GpC** indicated that van der Waals interactions would likely stabilize a densely-packed network, like the one schematized in Figure 95a. This dense network model presented unfavorable steric interactions and could not fit the experimental data whatsoever. Additionally, the G moieties display available H-bonding sites. If the molecules would stablish stabilizing H-bonds between the aminopyridine fragments or corners, as seen for various monomers in Chapter 2, the proposed cyclic trimers could form a flower- or star-like superstructure (Figure 95b,c). However, no pattern of the sort was observed in any of the STM images.

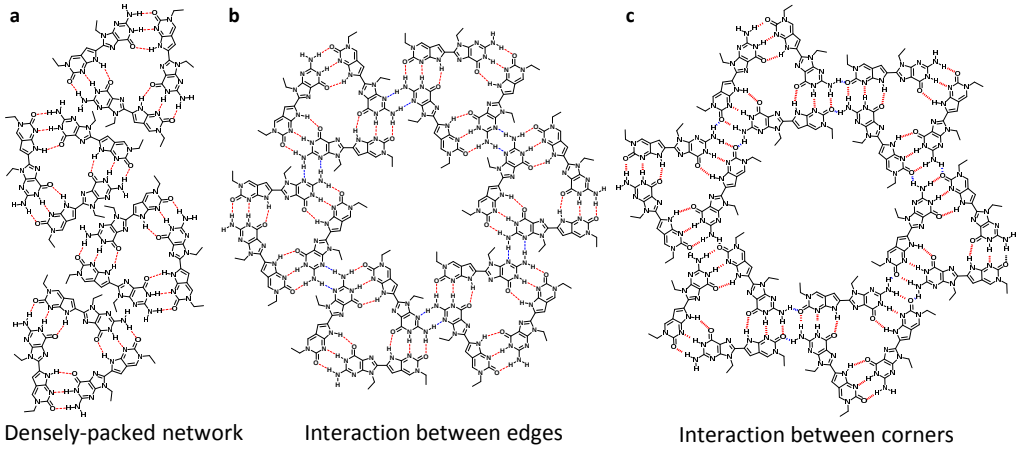


Figure 95. Proposals for the self-assembly of the cyclic trimers formed by GpC. (**a**) Densely-packed network of **GpC** cyclic trimers. (**b**) Flower-like superstructure formed by 6 **GpC** cyclic trimers, secondarily H-bonded through the G-aminopyridine fragments. (**c**) Star-like superstructure formed by six **GpC** cyclic trimers, secondarily H-bonded through their corners.

A first model supposed that the STM images corresponded to the adsorption of the **GpC** cyclic trimer (Figure 96b). The stabilization of the proposed assembly would mainly be driven by the establishment of 9 H-bonding interactions within the cyclic species. The network would be secondarily stabilized through van der Waals interactions between the alkyl tails. The experimental unit cell parameters (a = 2.0 nm; b = 2.0 nm; $\alpha = 60^{o}$) fitted this model, but as previously indicated by preliminary DFT calculations, van der Waals interactions would likely stabilize a more densely-packed network (Figure 95a). Additionally, this arrangement created two kind of pores that were not seen on the STM images: a first one corresponding to the inner cavity of the cyclic trimer; and a second one between macrocycles. Since this first proposal did not fit properly, other alternatives were considered.

As previously discussed, the STM images indicated the presence of a mismatched big bright dot (5 in Figure 88b) that could not be attributed to any molecule or molecular fragment. Because of this feature, another hypothesis contemplated the possibility of the stabilization of gold adatoms in the network.[244] Therefore, a second proposal was that three **GpC** molecules self-assembled into the formerly introduced H-bonded cyclic trimer and we proposed the co-adsorption of Au adatoms inside each of the macrocycles, as shown in Figure 96c. Let's remark that the three carbonyl groups of the G moieties point inwards the inner cavity of this trimer, which is thus electron rich. …

Another hypothesis, developed in Figure 96d and 96e was that three molecules formed a trimer, H-bonded through their G moieties. This disposition could work for both **GpC** and **GC7** and Au adatoms would be coadsorbed between three "pyrroloC" (Figure 96d) or cytosine moieties by stablishing three Au-O and three Au-N bonds (Figure 96e).[245] The shape of the formed cavity adapted well to the STM image. However, the calculated unit cell parameters were found to be too large (a_3 = 2.3 nm; b_3 = 2.3 nm; α_3 = 60º for **GpC** and a_4 = 2.4 nm; b_4 = 2.4 nm; α_4 = 60º for **GC7**) and did not verify the experimental data.

Further theoretical calculations are being carried out in order to support or dismiss our preliminary hypotheses and clarify what could be occurring when sublimating **GC7**.

Since monomer **GC7** did not afford the desired stable porous network formed by cyclic tetramers in these conditions, other candidates were examined for the self-assembly study under UHV.

[244] G. Pawin, K. L. Wong, D. Kim, D. Sun, L. Bartels, S. Hong, T. S. Rahman, R. Carp, M. Marsella, *Angew. Chem. Int. Ed.* **2008**, *47*, 8442–8445.

[245] a) P. Maksymovych, D. C. Sorescu, J. T. Yates, Jr., *Phys. Rev. Lett.* **2006**, *97*, 146103; b) P. Maksymovych, J. T. Yates, Jr., *J. Am. Chem. Soc.* **2008**, *130*, 7518–7519; c) Z. Shi, N. Lin, *J. Am. Chem. Soc.* **2009**, *131*, 5376–5377; d) Z. Yang, M. Corso, R. Robles, C. Lotze, R. Fitzner, A. Mena-Osteritz, P. Bäuerle, K. J. Franke, J. I. Pascual, *ACS Nano* **2014**, *8*, 10715–10722.

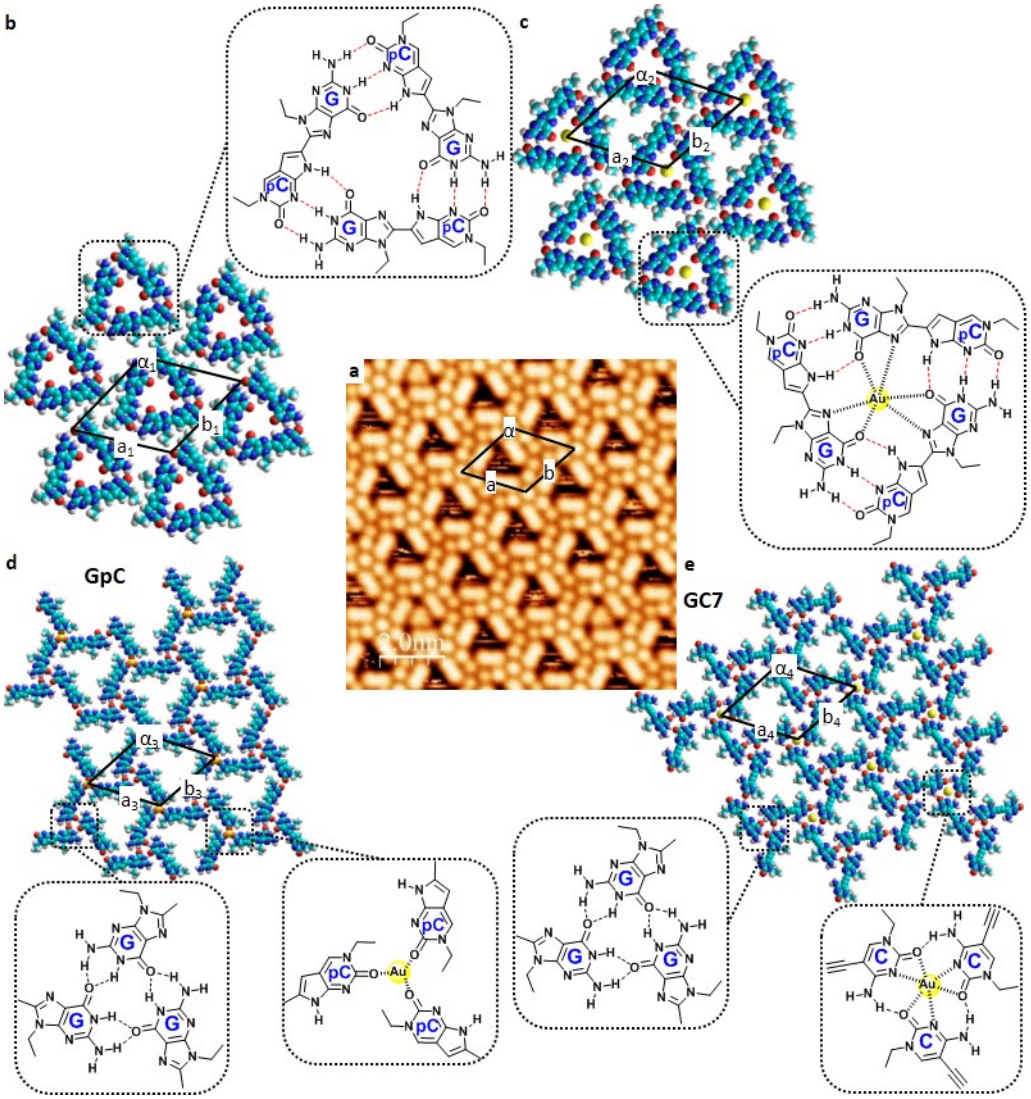
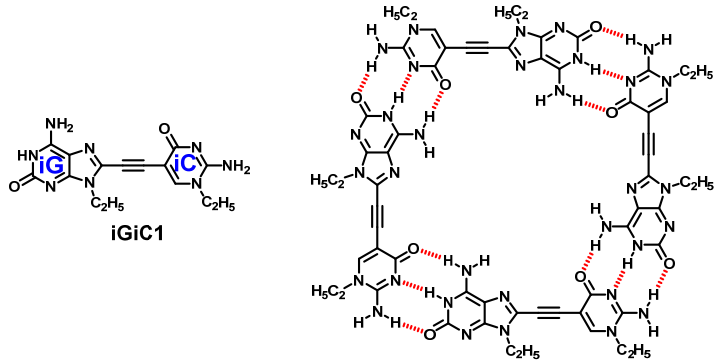


Figure 96. Different model proposals for the self-assembly observed at the Au (111)-vacuum interface. (**a**) Recorded UHV-STM image (V_t = 0.3 V; I_t = 10 pA). The unit cell is marked by black lines (a = 2.0 nm; b = 2.0 nm; α = 60º). (**b**) First proposed model of the network along with the chemical structure of the primary stabilizing motif corresponding to a H-bonded **GpC** cyclic trimer. Unit cell parameters: a_1 = 2.0 nm; b_1 = 2.0 nm; α_1 = 60º. (**c**) Second molecular model considering an H-bonded cyclic trimer based on **GpC** and the stabilization of a Au adatom inside the inner cavity. Unit cell parameters: a_2 = 2.0 nm; b_2 = 2.0 nm; α_2 = 60º. Gold adatoms are represented by a yellow sphere. (**d**) Third model based on an open oligomeric network based on **GpC**, along with the chemical structure of the two proposed stabilizing motifs. Unit cell parameters: a_3 = 2.3 nm; b_3 = 2.3 nm; α_3 = 60º. (**e**) Fourth proposed model based on an open oligomeric network, this time formed by **GC7**, along with the chemical structure of the two proposed stabilizing motifs. Unit cell parameters: a_4 = 2.4 nm; b_4 = 2.4 nm; α_4 = 60º.

3.2.2. Molecule iGiC1.

Monomer **iGiC1** uses the non-natural pair of complementary nucleobases: *iso*guanine and *iso*cytosine. The interaction between the couple is based on a *DDA–AAD* H-bonding pattern with two stabilizing secondary interactions, where no reverse Watson-Crick binding is possible and the resulting angle is 90°. Therefore the iG–iC binding is strong and directional. In analogy to **GC7**, two short ethyl chains have been chosen as substituents for the directing moieties. This second molecular target possess a low molecular weight (MW = 340.4 g/mol), for sublimation purposes, and the potential ability to form cyclic tetramers through the molecular recognition of iG and iC nucleobases (Scheme 24). In this case, the cyclization side-reaction seen for **GC7** should be less favored, since the carbonyl and amino functionalities positions are exchanged in *iso*cytosine derivatives. The lone pairs of the oxygen atom in the carbonyl moiety could attack the electrophilic triple bond. However, since we believe that the aromaticity of the heterocycle would not be recovered, the intramolecular cyclization event would be more unlikely.



Scheme 24. Chemical structure of the synthesized monomer **iGiC1** and proposed cyclic tetramer assembly.

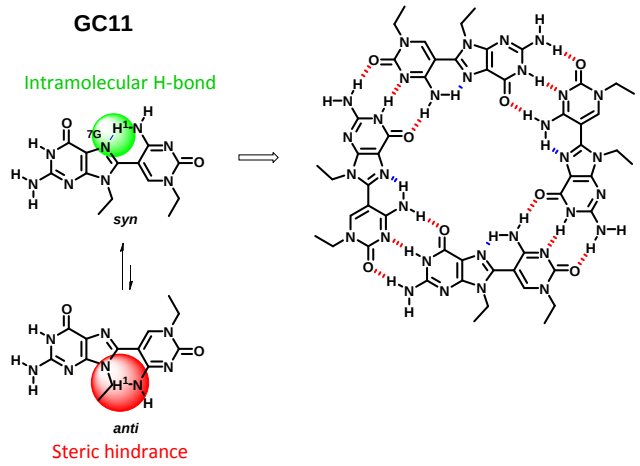
Monomer **iGiC1** will be investigated by UHV-STM in next experiments.

3.2.3. Molecule GC11.

Concerned by the reactivity of triple bonds on metallic surfaces during UHV-STM experiments,[246] molecule **GC11** was also targeted. This molecule, also based on the Watson-Crick GC base-pairing model, is the shortest possible monomer inspired by our common design: its nucleobases are solely separated by a simple single covalent bond.

[246] a) J. Björk, Y.-Q. Zhang, F. Klappenberger, J. V. Barth, S. Stafström, *J. Phys. Chem. C* **2014**. *118*, 3181–3187; b) Y.-Q. Zhang, N. Kepcija, M. Kleinschrodt, K. Diller, S. Fischer, A. C. Papageorgiou, F. Allegretti, J. Björk, S. Klyatskaya, F. Klappenberger, M. Ruben, J. V. Barth, *Nat. Commun.* **2012**, *3*, 1286; c) H. Zhou, J. Liu, S. Du, L. Zhang, G. Li, Y. Zhang, B. Z. Tang, H.-J. Gao, *J. Am. Chem. Soc.* **2014**, *136*, 5567–5570; d) V. K. Kanuru, G. Kyriakou, S. K. Beaumont, A. C. Papageorgiou, D. J. Watson, R. M. Lambert, *J. Am. Chem. Soc.* **2010**, *132*, 8081–8086; e) J. Liu, P. Ruffeux, X. Feng, K. Müllen, R. Fasel, *Chem. Commun.* **2014**, *50*, 11200–11203.

Besides the absence of the reactive triple bond, monomer **GC11** displays other interesting features. The rotation around this bond is influenced by the fact that the "*syn*" conformation would be favored by the formation of an intramolecular H-bond between H-1 of the C amino group and *N*-7 of the G-nucleobase, leading to a six-membered cycle. In contrast, the "*anti*" conformation would be discouraged by steric reasons (Scheme 25): the alkyl chain carried by the G-nucleobase and the C amino group are too close. In these terms, the free rotation of this bond is hindered and we would expect that the "*syn*" conformation should be largely favored compared to the opposite "*anti*" conformation. Let's note that the "*syn*" pattern allows the formation of the cyclic tetrameric species, whereas the "*anti*" favors the open-oligomeric architecture. Therefore, molecule **GC11** could potentially form the strongest tetrameric cyclic species, because of the elevated structural predisposition, resulting in high *EM* values.



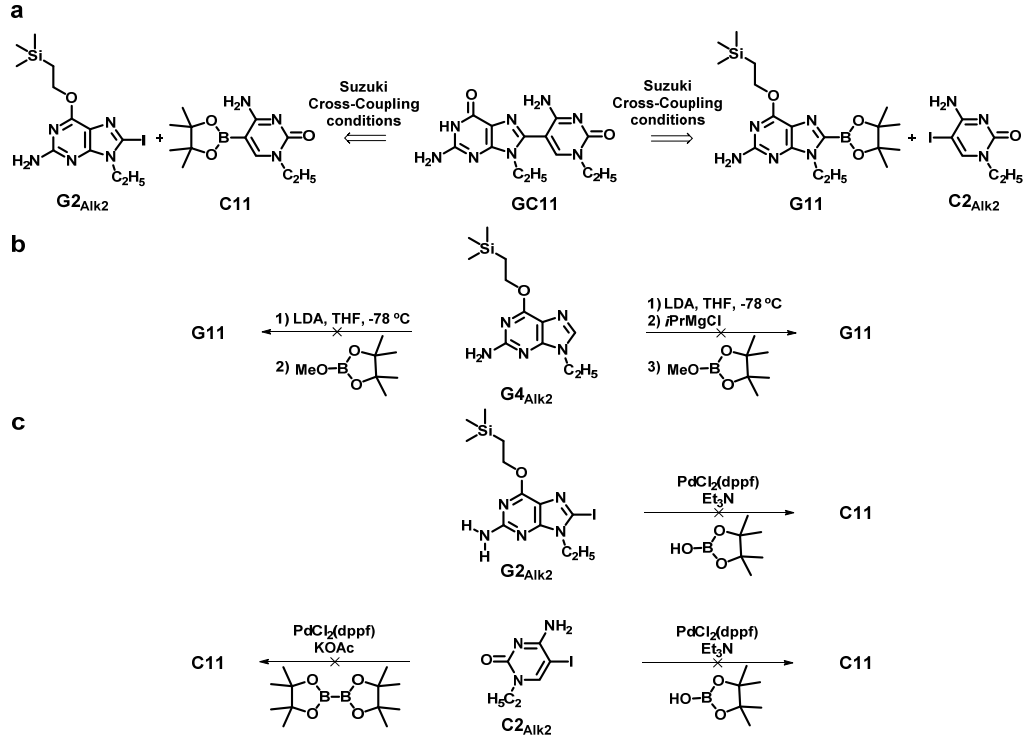
Scheme 25. Chemical structure of the targeted monomer **GC11**, "*syn*" and "*anti*" conformations equilibrium and proposed cyclic tetramer assembly.

For the preparation of the previous short GC monomers, the iodocytosine and the ethynyl-guanine derivatives were attached *via* a Sonogashira coupling reaction. In the synthetic pathway toward **GC11**, different strategies were considered (Scheme 26). The first retrosynthesis contemplates the coupling between the cytosine derivative $\mathbf{C2_{Alk2}}$ and a borylated-guanine derivative *via* a Suzuki Pd-catalyzed reaction (Scheme 26a, right). Another possibility would be to incorporate the borane derivative in the C-heterocycle and execute the Suzuki coupling reaction with $\mathbf{G2_{Alk2}}$ (Scheme 26a, left).

Deprotonation of the starting product $\mathbf{G4_{Alk2}}$ with freshly prepared LDA, followed by direct addition of methoxyboronic acid pinacol ester was first attempted for the borylation of the guanine derivative (Scheme 26b, left).[247] However, the borylation reaction did not occur under these conditions and mostly the starting material $\mathbf{G4_{Alk2}}$ was recovered. In a

[247] S. J. Mcquaker, C. L. Quinlan, S. T. Caldwell, M. D. Brand, R. C. Hartley, *ChemBioChem* **2013**, *14*, 993–1000.

second attempt, **G4$_{Alk2}$** was first deprotonated with LDA; then the Grignard reagent isopropylmagnesium chloride was added; and the resulting organomagnesium-guanine derivative was finally quenched with the dioxoborolane reagent, to afford product **G11** (Scheme 26b, right).[247] Unfortunately, once again only the starting material was recovered from a complex mixture.



Scheme 26. Possible synthetic routes toward **GC11**.

In view of these results, a Pd-catalyzed borylation strategy[248] was followed at this point. This protocol was first essayed on **G2$_{Alk2}$**, using pinacolborane in the presence of NEt_3 and $PdCl_2$(dppf),[249] but only the deiodination product was obtained (Scheme 26c, top). Then, we tested the reaction on the aryl iodide **C2$_{Alk2}$** both with pinacolborane[249] and bis(pinacolato)diboron in the presence of KOAc and $PdCl_2$(dppf) under strict anhydrous conditions (Scheme 26c, bottom).[250] Despite these are rather general conditions employed in Pd-catalyzed borylation of haloarenes, none of these attempts afforded the desired borylated product and only the dehalogenated derivatives could be isolated as the major product.

[248] W. Kin Chow, O. Ying Yuen, P. Ying Choy, C. Ming So, C. Po Lau, W. Tak Wong, F. Yee Kwong, *RSC Adv.* **2013**, *3*, 12518–12539.
[249] a) M. Murata, S. Watanabe, Y. Masuda, *J. Org. Chem.* **1997,** *62*, 6458–6459; b) M. Murata, T. Oyama, S. Watanabe, Yuzuru Masuda, **2000**, *65,* 164–168.
[250] T. Ishiyama, M. Murata, N. Miyaura, *J. Org. Chem.* **1995**, *60*, 7508–7510.

3.3. Summary and Conclusions.

The self-assembly process of a ditopic molecule, carrying complementary nucleobase derivatives on each side, has been investigated in this Chapter under vacuum and on different metallic substrates (Au (111) and Ag (111)). The experimental conditions for the sublimation of our relatively heavy and polar molecules were optimized for the deposition and imaging by STM under UHV. Molecule **GC7** was sublimated and deposited on Ag (111) and Au (111). Large domains of a well-resolved 2D porous network were imaged under UHV on the almost totally-covered gold surface. Numerous analyses, such as voltage evolution and annealing experiments, have been carried out in order to explore the self-assembled network. In this way, the first steps regarding the transfer of the supramolecular chemistry of our monomers, from the HOPG-liquid interface (Chapter 2) to the metal-vacuum interface, have been given. The experimental conditions used in Chapter 2 at the solid-liquid interface were more favorable for the self-assembly of our system, since they afforded a greater versatility and the role of dynamics was more relevant. Nonetheless, a very large and stable network was self-assembled on the surface using organic molecular beam epitaxy deposition.

A thorough analysis combining experimental manipulations, the construction of molecular models and theoretical calculations in the gas phase have been accomplished for understanding what it was occurring and gain insight into this self-assembly process. NMR spectroscopy, sublimation experiments in a cold-finger apparatus and theoretical calculations have been performed with the objective of elucidating the association mode on Au (111) when sublimating **GC7** at 200 ^{o}C. The main problem observed was that **GC7** suffered an intramolecular cyclization process induced by heating the solid sample. The product, **GpC**, may be the actual species being deposited onto Au (111) and this uncertainty complicated the analysis of the STM images. Further investigations will be carried out so as to get a better understanding of the self-assembly of this molecule on Au (111). Deposition at 4 K will be essayed, targeting the visualization of a single molecule. Additionally, non-contact AFM can be envisaged with the purpose of achieving better resolution and unequivocally build an appropriate molecular model corresponding to the observed STM images.

Three molecules have been synthesized and characterized in depth for their study by STM under UHV. A fourth molecule, **GC11**, comprising the two complementary nucleobases G and C solely attached by a single covalent bond, was attempted. A first retrosynthetic analysis considered the preparation of a borylated nucleobase derivative in order to carry out a Suzuki coupling reaction to yield **GC11**. Unfortunately, we failed to incorporate the pinacolborane moiety both on the G- and C-nucleobases. Alternatively, Stille cross-coupling reactions between **G2$_{Alk2}$** and the C-organostannane derivative, prepared with tributyltin chloride, could replace the Suzuki route. However, it has not been possible to carry out all the desired experiments during the lapse of this Thesis. Because of the unforeseen

synthetic difficulties, monomer **GC11** has not yet been synthesized but its preparation will be attempted shortly and **GC11**'s study by UHV-STM will be considered in further experiments. Monomers **iGiC1** and **GC4** will be also investigated by UHV-STM in the future. Additionally, the ability of the self-assembled 2D nanoporous networks to host appropriate guest molecules may be investigated further on.

3.4. Experimental Section.

3.4.1. General Methods.

The General Methods described in the previous Chapters are also applicable here.

Thermogravimetric analysis (TGA) analyses were performed under N_2 atmosphere in a TGA Q500 V20.13 Build 39 system from TA Instruments at the Instituto IMDEA Nanociencia.

Differential Scanning Calorimetry (DSC) experiments were carried out in a Discovery DSC calorimeter from TA Instruments in the SIdI at the UAM.

3.4.2. Experimental UHV-Setup and STM Measurements.

The STM experiments under UHV conditions described in this Chapter were carried out in an ultrahigh-vacuum (UHV) chamber with a base pressure of 5 x 10^{-11} mbar. The system was purchased from Omicron and is equipped with a low temperature scanning tunneling microscope (LT-STM), low energy electron diffraction (LEED), Auger electron spectroscopy (AES), sample and STM tip transfer and heating capabilities, several interchangeable evaporation cells, an ion gun for sample cleaning purposes, and a quadrupole mass spectrometer. It can be operated under He (4 K). However, N_2 (77 K) was used in these experiments. Tungsten tips were prepared by electrochemical etching. Following their introduction into the preparation chamber, tips were sputtered and annealed to remove superficial contaminants, including oxide layer coatings. Electrochemical etching of ultrasharp STM tips was carried out using KOH. Tips were then sputtered with Ar^+ ions (1.5 keV beam energy; ion current of 15 μA) and annealing flashes. Au (111) surfaces were prepared by successive cycles of sputtering with Ar^+ ions and annealing at 900 K until the herringbone (22x√3) reconstruction, characteristic of clean Au (111), was observed in STM images. **GC7** was first outgassed in UHV and then sublimated from a 3-cell UHV evaporator [KENTAX] on Au (111) at room temperature. The storage vessel can be pumped separately by the turbo pump to outgas the impurities. **GC7** was dosed into the preparation chamber until a given partial pressure was achieved.

All the surface manipulation experiments, data acquisition, and image processing were performed using the WSxM software.[251] STM images were acquired in constant-current mode at low temperature (105–120 K) using W tips.

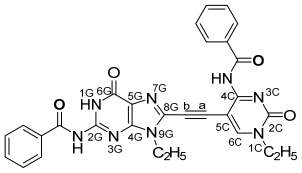
[251] I. Horcas, R. Fernández, J. M. Gómez-Rodríguez, J. Colchero, J. Gómez-Herrero, A. M. Baró, *Rev. Sci. Instrum.* **2007**, *78*, 013705.

3.4.3. Computational Details.

First-principles calculations have been performed using the Vienna *ab initio* simulation package (VASP).[252] The generalized gradient approximation (GGA) for the exchange-correlation functional of Perdew, Burke, and Ernzerhof (PBE) is employed.[253] To describe the effective potential of core electrons, the projector-augmented wave (PAW) method[254] is utilized. A cutoff energy of 450 eV for the plane wave expansion and a Methfessel-Paxton smearing[255] of 0.2 eV for the occupation of electronic states are chosen. For the monomers, the Brillouin-zone sampling is performed with a 3 x 3 x 3 k-point mesh according to the Monkhorst-Pack grid.[256] For the self-assembled networks, a 3 x 3 x 1 k-point mesh is used. The molecules or molecular networks are placed in a 3D periodic cell, with ~10 Å of vacuum in the directions where the interaction between adjacent unit cells should be avoided. During the geometry optimizations, all atomic positions are fully relaxed. The local minimum of potential energy is searched by a conjugate-gradient algorithm until all forces on each ion are smaller than 0.02 eV.Å^{-1}. The lattice parameters of the molecular networks have also been optimized with an accuracy of 0.1 Å.

3.4.4. Synthesis and Characterization.

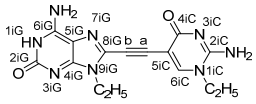
The synthesis and characterization of monomers **GC4** and **GC7** was described in the previous Chapter 2 (section **2.4.2.**).



GC7p. **GC7** (36 mg, 0.11 mmol) was suspended in pyridine (1.5 mL) and cooled down to 0 ^{o}C. Then, benzyl chloride (30 µL, 0.25 mmol) was added dropwise. The resulting clear red solution was let to reach rt. The reaction was completed in 6 h. The reaction mixture was quenched with water (4 mL) and stirred for 20 min. NH_3 (aq) (1 mL) was added afterwards and the mixture was stirred at rt for 10 min. The crude product was extracted with CH_2Cl_2 (3 × 20 mL). The combined organic extracts were washed with $NaHCO_3$ (5%) (1 × 20 mL) and water (1 × 20 mL), dried over $MgSO_4$ and evaporated. The resulting solid was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1), affording 6.4 mg of **GC7p** (11% yield) as a pale solid.

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 13.29 (bs, 1H, $N^{1G}H$), 12.20 (bs, 1H, $N^{2G}H$), 8.98 (bs, 1H, $C^{4C}NH$), 8.28 (d, *J* = 7.4 Hz, 2H, Ar), 7.97 (d, *J* = 7.4 Hz, 2H, Ar), 7.96 (s, 1H, H^{6C}), 7.56 (m, 5H, Ar), 7.42 (t, *J* = 7.5 Hz, 1H, H^{p}), 4.30 (q, *J* = 7.0 Hz, 2H, $N^{9G}CH_2CH_3$), 3.93 (q, *J* = 7.0 Hz, 2H, $N^{1C}CH_2CH_3$), 1.42 (m, 6H, $N^{9G}CH_2CH_3$, $N^{1C}CH_2CH_3$).

[252] a) G. Kresse, J. Hafner, *Phys. Rev. B* **1994**, *49*, 14251–14269; b) G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169–11186; c) G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **1996**, *6*, 15–50.
[253] J. P. Perdew, W. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
[254] a) P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953–17979; b) G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758–1775.
[255] M. Methfessel, A. Paxton, *Phys. Rev. B* **1989**, *40*, 3616–3621.
[256] H. J. Monkhorst, J. D. Pack, *Phys. Rev. B* **1976**, *13*, 5188–5192.



iGiC1. A dry DMF/NEt_3 (4:1) solvent mixture (20 mL) was subjected to deoxygenation by three freeze-pump-thaw cycles with argon. Then, this solvent was added over the system containing **iG1$_{Alk2}$** (0.49 mmol, 0.10 g), **iC2$_{Alk2}$** (1.49 mmol, 0.39 g), CuI (0.005 mmol, 0.9 mg) and $Pd(PPh_3)_4$ (0.010 mmol, 11.5 mg). The reaction was stirred under argon at 40 ^{o}C overnight until completion, which was monitored by TLC. The resulting insoluble dark solid in suspension was filtered and washed with THF, acetonitrile and DMF, affording **iGiC1** as a brown solid (80 mg, 47%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.00 (s, 1H, H^{6iC}), 7.83 (bs, 2H, $C^{2iC}NH_2$), 4.01 (q, J = 7.2 Hz, 2H, $N^{1iC}CH_2CH_3$), 3.93 (q, J = 7.2 Hz, 2H, $N^{9iG}CH_2CH_3$), 1.28 (t, J = 7.2 Hz, 3H, $N^{9iG}CH_2CH_3$), 1.13 (t, J = 7.2 Hz, 3H, $N^{1iC}CH_2CH_3$).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 152.4, 148.2, 142.6, 131.7, 131.6, 129.9, 128.9, 128.8, 110.9, 79.5, 79.0, 77.2, 72.5, 14.8.

HRMS (ESI+): Calculated for $C_{15}H_{17}N_8O_2$ $[M+H]^+$: 341.1468. Found: 341.1479.

Chapter 4.

Preparation and Characterization of H-bonded Nanoporous Langmuir-Blodgett Films.

In Chapter 4 the preparation of monolayers based on our molecular dinucleobase design, this time at the water-air interface, and their transfer by the vertical deposition method onto solid substrates will be attempted. The fabrication of monolayers at the air-liquid interface using the Langmuir-Blodgett technique, described in this Chapter, as well as the characterization of these films by AFM and STM at the solid-air interface at room temperature are being carried out at the *Instituto de Nanociencia de Aragón* (Zaragoza) in collaboration with the group of Dr. Pilar Cea and Dr. Santiago Martín. The preliminary results obtained so far are exposed in this Chapter.

4.1. Introduction. The Langmuir-Blodgett technique.

The Langmuir-Blodgett (LB) technique owes its name to Irving Langmuir (1881-1957) and Katharine Blodgett (1898-1979) who developed the fabrication of organic films on the surface of a liquid and their transfer onto a solid substrate. Irving Langmuir showed that after spreading fatty acids, alcohols or esters on the water surface and reducing the available area, all the molecules acquired the same orientation and became organized, forming the so-called "Langmuir film", whose thickness corresponded to the length of a single molecule.[257] Later, Katharine Blodgett studied under Langmuir supervision the possibility of transferring such films onto solid substrates by successive deposition of mono-molecular layers.[258] She showed that when a substrate, placed perpendicular to the film on the aqueous surface, is slowly immersed into and withdrawn from the subphase, mono- to multi-layered films can be obtained, depending on the number of immersion and emersion cycles. They were called "Langmuir-Blodgett films" or "LB films".

One of the most important features of the LB method is the large variety of substrates that can be employed including metals, carbon, conducting polymers, silicon, graphene-based electrodes, etc. However, the real breakthrough in the study of LB films was due to Hans Kuhn,[259] who in the 1970s synthetized and studied molecules specially designed for the construction of organized molecular buildings. Since then, many researchers have used the LB technique for the assembly of organic and hybrid organic-inorganic compounds into well-ordered films for different fields of applications. In fact, the LB technique is nowadays one of the most sophisticated bottom-up methods for the assembly of molecular materials on surfaces.[260]

A Langmuir-Blodgett trough (Figure 97) is used for LB films fabrication and subsequent transfer onto a solid substrate. It has four main components: (1) a container where the subphase, generally water, is placed. (2) A mobile barrier (or two) that controls

[257] I. Langmuir, *J. Am. Chem. Soc.* **1916**, *38*, 2221–2295.
[258] K. B. Blodgett, *J. Am. Chem. Soc.* **1935**, *57*, 1007–1022.
[259] H. Kuhn, *J. Chem. Phys.* **1970**, *53*, 101–108.
[260] a) A. Ulman, *An Introduction to ultrathin organic films: from Langmuir-Blodgett to self-assembly*, Academic Press, Inc. San Diego, **1991**; b) G. L. Gaines, *Insoluble monolayers at liquid-gas interface*, Interscience, John Wiley & Sons, New York, **1966**; c) M. C. Petty, *Langmuir-Blodgett films: an introduction*, Cambridge University Press, Cambridge, **1996**.

the area available for the molecules on the subphase. Both the trough and the barrier are made out of an inert and hydrophobic material, such as Teflon™, so that the subphase would form a convex meniscus with the walls of the container and the mobile barrier, and to avoid any possible contamination. (3) A surface tension sensor, usually a Wilhelmy plate, for monitoring the LB film formation. (4) A dipping mechanism that holds the solid substrate and enables controlled deposition cycles.

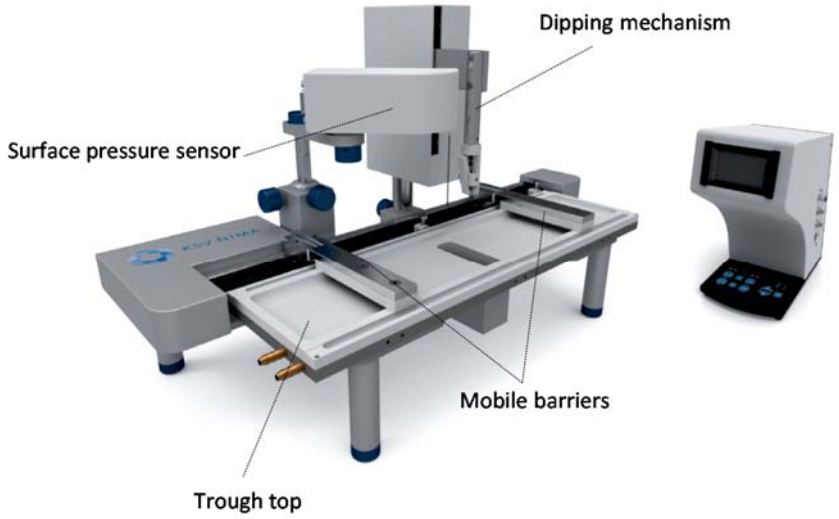


Figure 97. KSV NIMA Langmuir-Blodgett Trough (Model KN 2002).[261]

Figure 98 illustrates the fundamentals of the LB technique. In the LB method, the material is first ordered at the air-liquid interface and then deposited onto the substrate in contrast to most self-assembly methods where the head group of the molecule is first chemisorbed to the substrate and the molecules are ordered later in 2D. The nature of the substrate-molecule interaction depends on both the nature of the surface and the functional group of the organic material. Thus, the monolayer can be either physisorbed onto the substrate surface or chemisorbed.[262] Therefore, by making use of the large variety of functional groups that can be physically or chemically adsorbed onto different electrodes, it is possible to use the LB technique to fabricate molecular junctions featuring anyone of a large number of organic-electrode interfaces.[263] The LB method involves a detailed and relatively long experimental procedure which precludes its use as a manufacturing technique. However, it is an excellent scientific tool with a tremendous flexibility. This is the reason why it has been widely used in the literature for the fabrication of molecular junctions.

[261] www.biolinscientific.com
[262] a) S.Gyepi-Garbrah, R. Šilerova, *Phys. Chem. Chem. Phys.* **2002**, *4*, 3436–3442; b) P. Cea, M. C. Lopez, S. Martin, A. Villares, G. Pera, I. Giner, *J. Chem. Edu.* **2009**, *86*, 723–725.
[263] G. Pera, A. Villares, M. C. Lopez, P. Cea, D.P. Lydon, P.J. Low, *Chem. Mater.* **2007**, *19*, 857–864.

The technique for the preparation of LB films is divided in three steps: the spreading, the compression and the transference processes.[264] An organic solution is deposited dropwise on the subphase surface during the spreading process (Figure 98a). Careful deposition close to the liquid surface, covering all the available surface and keeping the surface pressure close to 0 mN/m is needed. Volume and concentration of the spreaded organic solution are fixed and therefore, the number of deposited molecules is known. Low boiling point solvents that are not miscible with water are used, such as chloroform, for fast evaporation (*ca.* 15 min). After a complete evaporation of the solvent, the compression process starts by means of the mobile barrier (Figure 98b). Upon compression, a surface pressure *vs* area per molecule isotherm is registered *in situ* (Figure 98c). This isotherm provides information about the phases and phase transitions of the monolayer. The molecules go through distinct phases (gas, liquid expanded, liquid condensed) depending on the interactions between them, until they form a compact and homogenous monolayer (solid phase) (Figure 98d). If the compression process were to be continued, after the solid phase had been reached, the monolayer would collapse due to an excess of the minimum value of physical area occupied by the molecule.

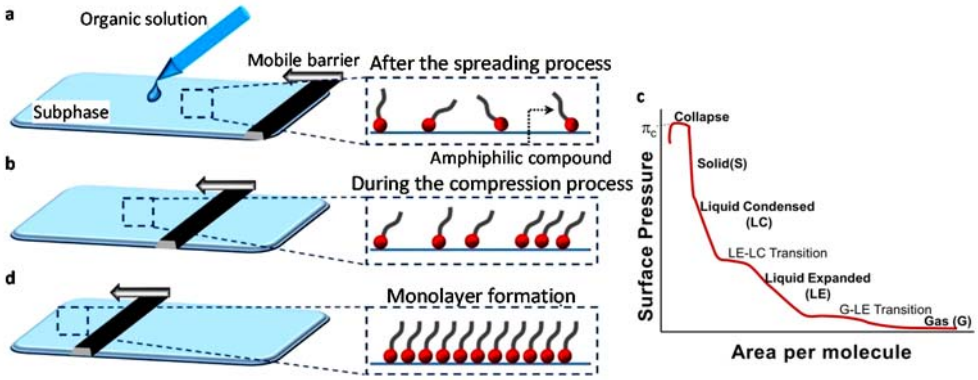


Figure 98. Schematic representation of the Langmuir-Blodgett method.[264]

Once a stable monolayer has been formed at the air-liquid interface, the Langmuir film can be transferred onto a solid support. If the substrate is initially out of the subphase, the transference is mediated by hydrophobic interactions between the molecules and the substrate (Figure 99a). However, if the substrate is initially immersed in the liquid phase, the transference is mediated by molecule/substrate hydrophilic interactions (Figure 99b).[265] Alternatively, very rigid monolayers can be transferred by a parallel contact of the solid substrate with the liquid subphase containing the Langmuir film (horizontal deposition) to yield Langmuir-Schaefer (LS) films.[266] In this case, the monolayer/substrate

[264] P. Cea, L. M. Ballesteros, S. Martín, *Nanofabrication* **2014**, *1*, 96–117.
[265] L. M. Ballesteros, S. Martín, G. Pera, P. A. Schauer, N. J. Kay, M. C. López, P. J. Low, R. J. Nichols, P. Cea, *Langmuir* **2011**; *27*, 3600–3610.
[266] I. Langmuir, V. J. Schaefer, *J. Am. Chem. Soc.* **1938**, *60*, 1351–1360.

union has a hydrophobic nature (Figure 99c). After repeated immersion and emersion passages of the substrate through the monolayer at the air-liquid interface, ordered mono- to multi-layered films, whose thickness depends on the number of cycles, can be achieved. Type Z, X or Y films can be reached (Figure 99d). In Y-type deposition the monolayer is transferred on both the emersion and the immersion of the substrate. Centrosymmetric multilayers are thus obtained, where the interactions are head (hydrophilic group)-to-head and tail (hydrophobic chain)-to-tail. On the contrary, X and Z-type films refer to situations where deposition only occurs on the immersion or on the emersion, respectively. No centrosymmetric films are recovered instead, where head-to-tail (X-type transference) and tail-to head (Z-type) interactions are involved.

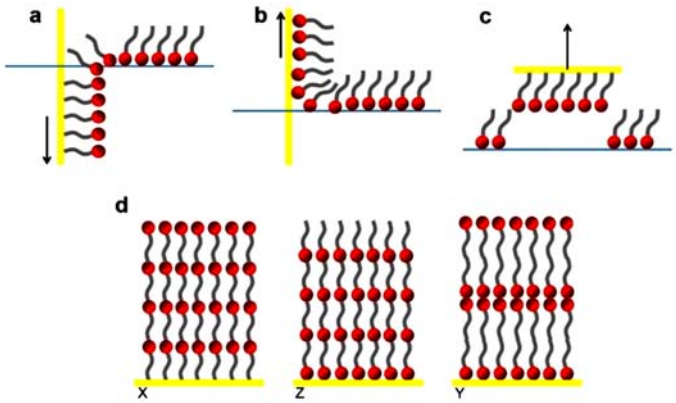
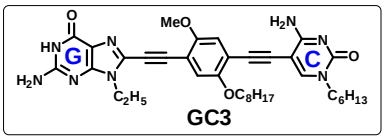


Figure 99. Langmuir-Blodgett technique for the transference of a monolayer when (**a**) the substrate is initially out of the subphase; (**b**) in the subphase; and (**c**) *via* the Langmuir-Schaefer method. (**d**) Possible LB films.

Our contribution:

In this Chapter, we aim to reproduce the results previously obtained with our molecules onto HOPG at the solid-liquid interface (see Chapter 2). We would like to broaden the investigation of the self-assembly of our monomers to the air-liquid interface and profit from the LB technique to evaluate the preparation of multilayered nanoporous systems in a controlled manner. Monomer **GC3** has been selected for these studies (Scheme 27) because it formed large and very stable H-bonded 2D networks of cyclic tetramers (see Chapter 2, section **2.2.1.**). Moreover, this complementary GC unsymmetric monomer was proven able to host suitable guest molecules within its tailored nanocavities, endowing the LB films with supplementary applications.



Scheme 27. Considered target monomer in Chapter 4 for the generation of LB films.

4.2. Results and Discussion.

This work has been divided in two main parts, depending on the considered substrate for vertical deposition: HOPG or Au (111). Since our self-assembled monolayers had been previously studied mostly on HOPG (see Chapter 2), this substrate was the first one considered.

4.2.1. LB films onto HOPG.

Most effort was dedicated at the beginning to optimize the experimental conditions for the preparation of adequate Langmuir films at the water-air interface, previous to the transference process onto solid substrates. Molecule **GC3** solely dissolved in chloroform did not present surface activity on the water surface and therefore, Langmuir films could not be prepared. However, by adding ethanol adequate Langmuir films could be obtained. A 5 x 10^{-6} M solution of **GC3** in chloroform was prepared and diluted to 4 x 10^{-6} M by adding EtOH, thus obtaining a 4:1 $CHCl_3$:EtOH mixture. The preparation of a Langmuir film was attempted by spreading 920 µL of this solution of **GC3** in a 4:1 $CHCl_3$:EtOH mixture onto the subphase. The solvent was let to evaporate for several minutes and thereafter, the compression process began by moving the mobile barrier, forming a monolayer at the water-air interface and the surface pressure isotherm was registered *in situ* (Figure 100). Let's remark that not all the molecules display a surface pressure isotherm where all the phases and transitions can be appreciated, like in the generic isotherm showed in Figure 98c.

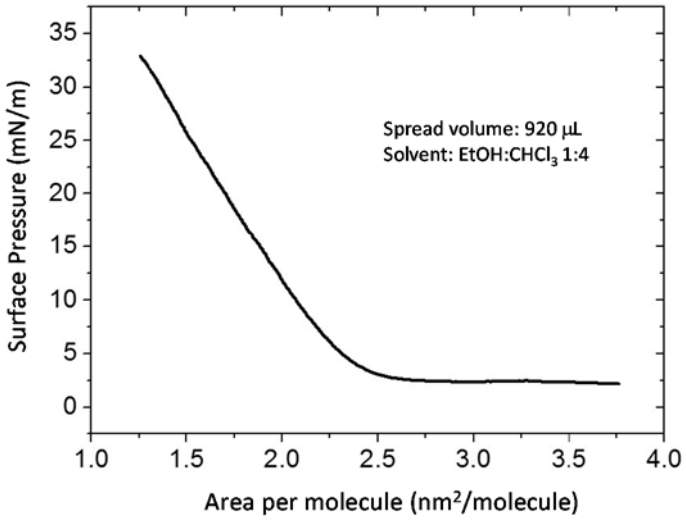


Figure 100. Surface pressure isotherm of a LB film transferred onto HOPG at 30 mN/m from a monolayer prepared by spreading 920 µL of a 4 µM **GC3** solution in EtOH:$CHCl_3$ (1:4).

EtOH was proven fundamental for the exhibition of surface activity and the subsequent preparation of Langmuir films. Large **GC3** aggregates are believed to be formed in $CHCl_3$, in view of the poor solubility of this compound in this solvent. EtOH might prevent their formation, permitting a good spreading of the molecule onto the water surface. EtOH probably stablishes H-bonding with the **GC3** molecules and partially breaks their aggregation. The surface pressure isotherm was recorded *in situ* during the compression process and showed a significant increase of the surface pressure at a value of ~2.5 nm^2/molecule. The self-assembled Langmuir film at the water-air interface was then transferred onto HOPG at 30 mN/m. The substrate was initially immersed in the subphase and then pulled out of it. This transference would correspond to the situation described in (Figure 99b). Let's remark that in the STM images of the nanoporous network formed by **GC3** onto HOPG at the solid-liquid interface (see Figure 57 in Chapter 2, section **2.2.1.**), the area *per* molecule was measured as 3.2 nm^2/molecule. This can be understood as the theoretical area *per* molecule value when all the **GC3** molecules are self-assembled in 2D forming a network of cyclic tetramers, bonded through their aminopyridine fragments. When comparing these two values, one can see that in the transferred film formed at the water-air interface, the molecules are more densely packed. This could correlate to the molecules reaching out of the monolayer plane in some areas of the film or to the partial formation of multilayers. The morphology of the resulting film was finally analyzed by AFM (Figure 101).

AFM images showed the formation of graphene-like flake structures, randomly distributed on the basal plane of HOPG (Figure 101a). These flakes ranged 1–5 nm high and 50–300 nm wide. Nucleation seemed to start at HOPG steps and defects (decorated steps) and thereafter the structure growth continued to the HOPG terraces leaving free zones of the substrate. Cross-section analysis showed the presence of layers growing from HOPG steps whose minimum height (0.3 nm) was compatible with a **GC3** monolayer assembled by π-stacking onto the basal plane of HOPG (red arrows) (Figure 101b). The images showed 1–2 nm high flakes, marked with blue arrows, that could be estimated as a few monolayers (considering an interlayer separation of 0.3 nm) and also higher flakes. The adhesion histogram (turquoise line in Figure 101c) could be deconvoluted into three different contributions (red, black and blue lines), corresponding to three different type of interactions with the AFM tip (from the substrate, from the flakes and from the different arrangements of the **GC3** molecules). Adhesion AFM images (Figure 101d) showed as well these distinct **GC3** organizations by exposing a differential chemical composition when interacting with the AFM tip. However, the high-flake form on HOPG resulted to be highly non-conducting, exhibiting a band gap of almost 5 V, obtained from STS data (Figure 101e).

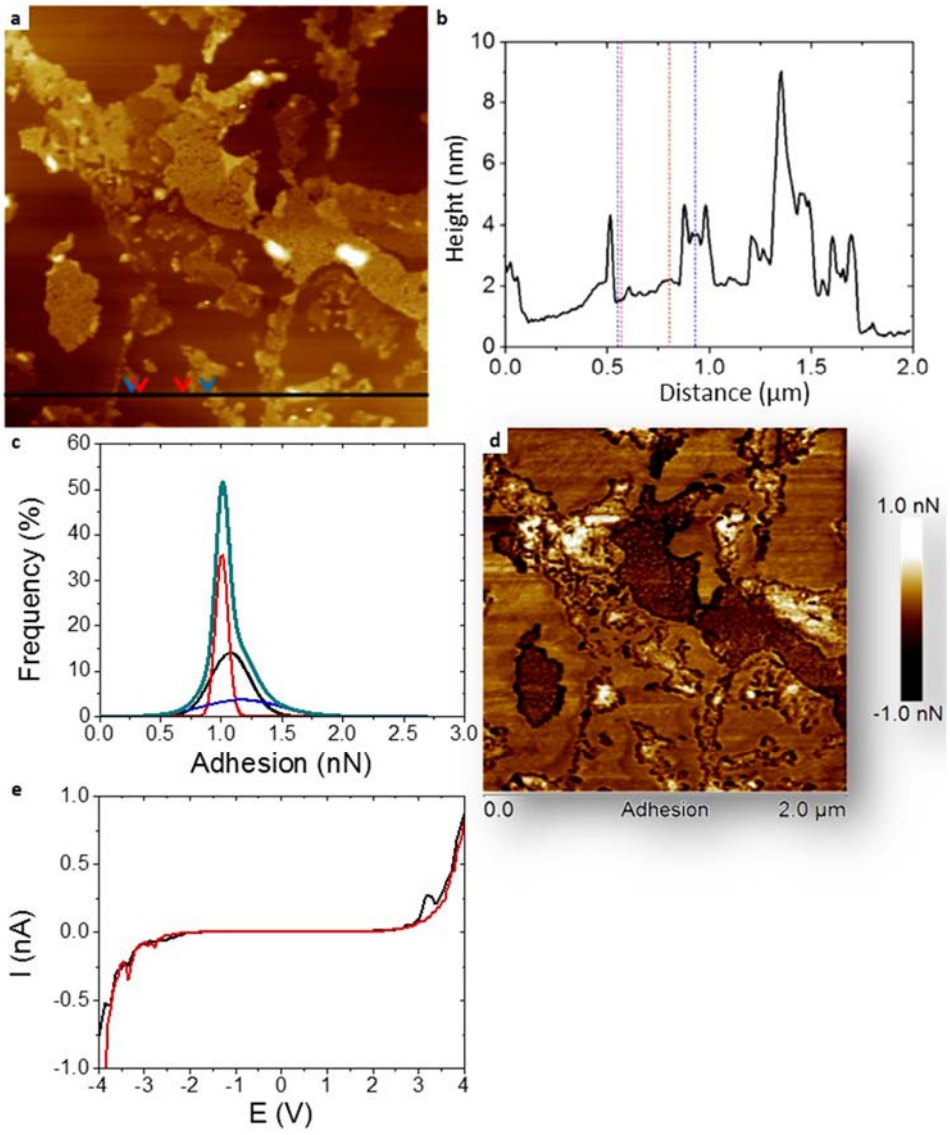


Figure 101. Characterization of a LB film transferred onto HOPG at 30 mN/m from a monolayer prepared by spreading 920 µL of a GC3 4 µM solution in EtOH:$CHCl_3$ (1:4). (**a**) 2.0 x 2.0 μm^2 topographic AFM image. (**b**) Typical cross-section analysis of the **GC3** aggregates on HOPG. (**c**) Histogram showing the adhesion force distribution. (**d**) Corresponding adhesion image of the AFM image showed in (**a**). (**e**) Typical STS IV curve obtained for the **GC3** aggregates on HOPG.

By using the transference conditions previously indicated (30 mN/m), a small area seemed to be covered by a single monolayer but the presence of multilayers (three or more) seemed to be dominant. Therefore, other transference experimental conditions were considered (Figure 102).

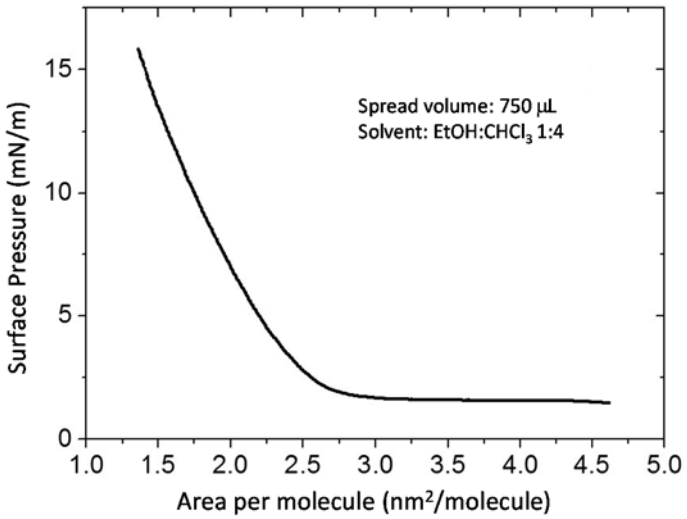


Figure 102. Surface pressure isotherm of a LB film transferred onto HOPG at 15 mN/m from a monolayer prepared by spreading 750 µL of a 4 µM **GC3** solution in EtOH:$CHCl_3$ (1:4).

When transferring the Langmuir films at a lower surface pressure, such as 15 mN/m, and analyzing by AFM the morphology of the film, a less homogeneous surface was appreciated (Figure 103). On the one hand, areas with a similar morphology to the one observed for the former sample (Figure 101) were encountered. Cross-section analysis (Figure 103b) revealed again the presence of flakes but now with a lower height, not surpassing 1–1.5 nm (marked with red arrows). More interestingly, other HOPG terrace zones, characterized by the presence of an incomplete monolayer (0.33 nm high, red arrow), and even a bilayer (0.67 nm high, blue arrow), decorated with some nanoparticles ranging 0.5–1.5 nm, could also be detected (Figure 103c,d). The quantitative adhesion map (Figure 103f) revealed a differential interaction between the HOPG free surface, the **GC3** monolayer/bilayer and the 3D nanoparticles on top.

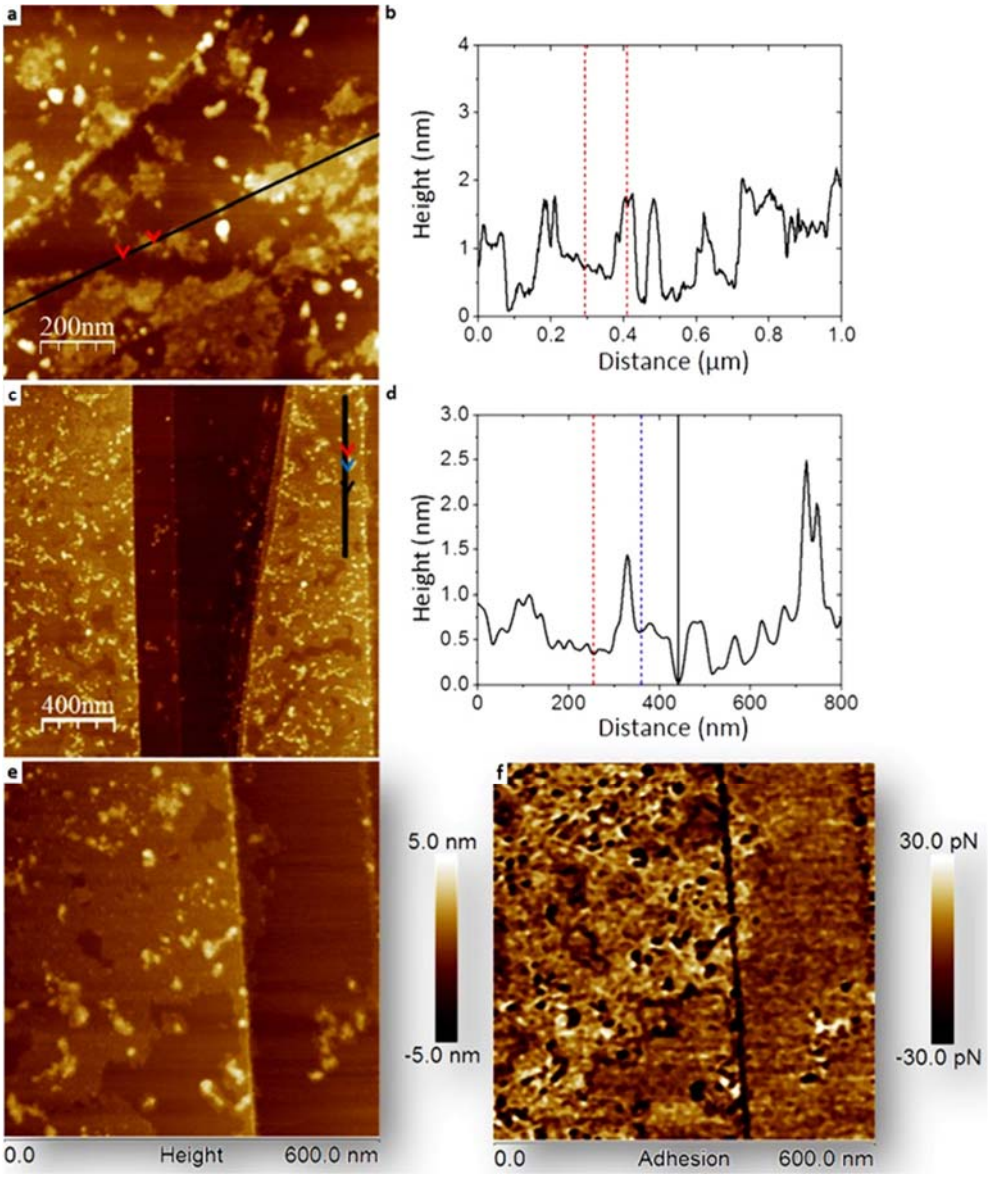


Figure 103. Characterization of a LB film transferred onto HOPG at 15 mN/m from a monolayer prepared by spreading 750 µL of a GC3 4 µM solution in EtOH:$CHCl_3$ (1:4). (**a**, **c**) Topographic AFM images. (**b**,**d**) Corresponding typical cross-section analyses of the **GC3** aggregates on HOPG. (**e**) 600 x 600 nm^2 topographic AFM image and (**f**) its corresponding adhesion image.

4.2.2. LB films onto Au (111).

At this point, the vertical deposition onto the gold substrate and subsequent characterization of the LB film by AFM was attempted (Figure 104).

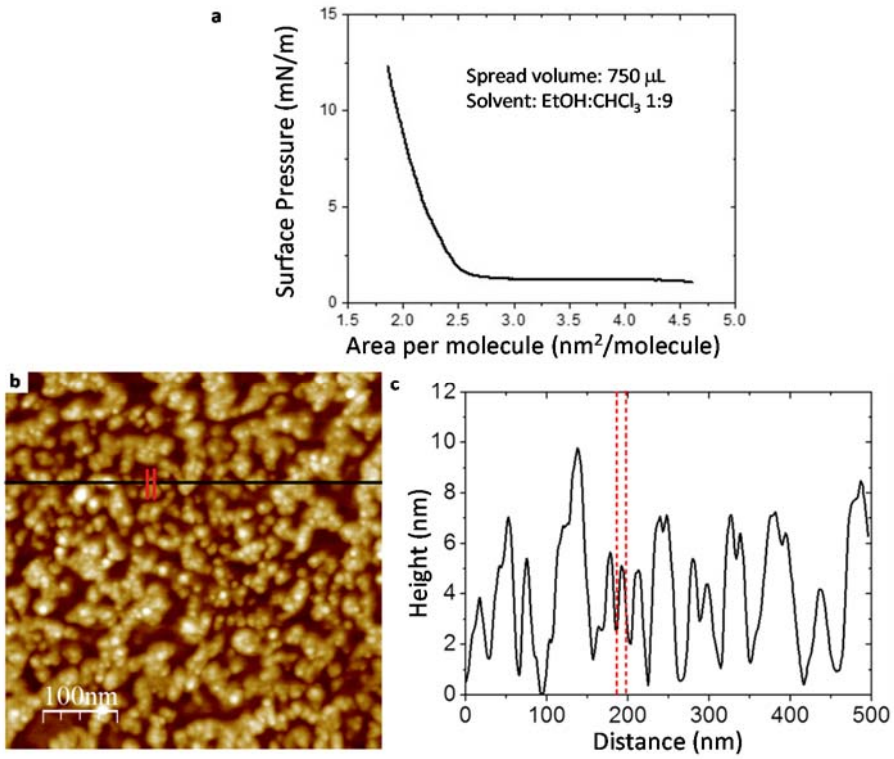


Figure 104. Characterization of a LB film transferred onto Au (111) at 12 mN/m from a monolayer prepared by spreading 750 µL of a GC3 4 µM solution in EtOH:$CHCl_3$ (1:9). (**a**) Surface pressure isotherm. (**b**) Topographic AFM image. (**c**) Corresponding typical cross-section analysis of the **GC3** aggregates on Au (111).

A first experiment was carried out where 750 µL of a **GC3** solution in a 9:1 $CHCl_3$:EtOH mixture were spread onto the subphase. The mobile barriers started to move so as to form the Langmuir film, which was then transferred onto the gold substrate at 12 mN/m. The higher ratio of chloroform resulted in an increase of the aggregation and no formation of monolayers or bilayers was observed. AFM images showed a completely different morphology of the film on Au (111), compared to HOPG, with the formation of branched islands formed by 3D-aggregates (Figure 104b). The **GC3** nanoparticle islands were measured to be 3–7 nm high and 20–150 nm wide. On the other hand, the individual nanoparticles ranged 1–3 nm high and 8–15 nm wide. It is worth mentioning that similar problems related to the aggregation and the insolubility of the products were previously

observed. In Chapter 2, the addition of a co-solvent, such as octanoic acid, was proven necessary to break the aggregation, solubilize the sample and obtain adequate images by STM at the solid-liquid interface. In the following experiments, the amount of EtOH for the formation of Langmuir films at the water-air interface was increased to match the ratio used in the deposition experiments onto HOPG (Figure 105).

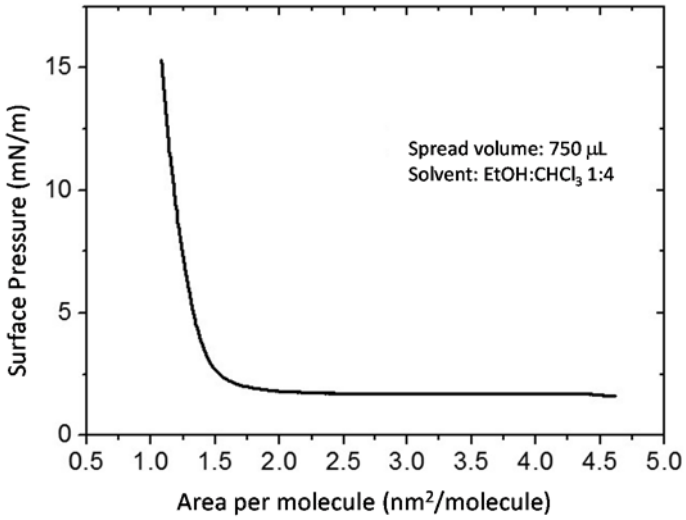


Figure 105. Surface pressure isotherm of a LB film transferred onto Au (111) at 15 mN/m from a monolayer prepared by spreading 750 µL of a 4 µM **GC3** solution in EtOH:$CHCl_3$ (1:4).

This time the surface pressure isotherm displayed a different behavior of the film and an acute increase of the surface pressure was identified for a smaller area when spreading a small volume of the 4 µM **GC3** solution. The resulting transferred film was analyzed by AFM (Figure 106).

This time, the presence of a monolayer (0.33 nm high, green arrows) and bilayer (0.66 nm high, red arrows) with porous features of 2–5 nm diameter (measured horizontal distance), homogenously arranged onto the substrate, could be more clearly detected in the AFM images (Figure 106a,b). On top of this, 1 nm high flakes (blue arrow), decorated with some nanoparticles ranging 0.5–1.5 nm high and 8–15 nm wide, were also seen. The quantitative adhesion map (Figure 106d) revealed a differential interaction between the Au (111) free surface, the **GC3** monolayer/bilayer and the 1 nm high flakes.

The LB film was also studied by STM and the images of the adsorbed layer showed now no trace of the flakes observed by AFM, indicating thus their poor conductivity. Nevertheless, an ultrathin porous-shaped organic monolayer was noticeable. The presence of a porous network on the Au (111) surface was seen by power spectral density analysis with a maximum at 2.31 nm and also a significant major shoulder at 1.93 nm (Figure 106f).

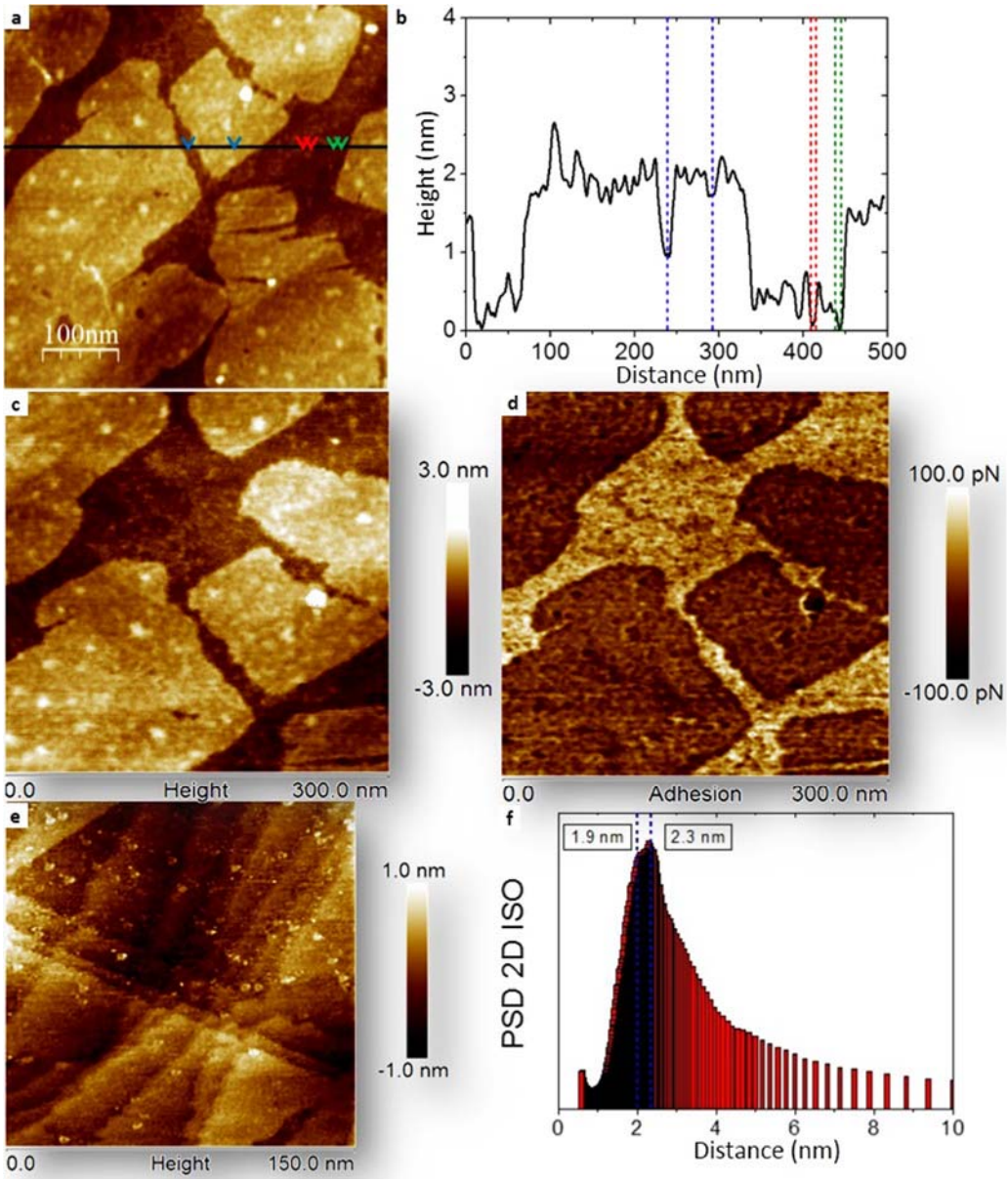


Figure 106. Characterization of a LB film transferred onto Au (111) at 15 mN/m from a monolayer prepared by spreading 750 µL of a GC3 4 µM solution in EtOH:$CHCl_3$ (1:4). (**a**) Topographic AFM image. (**b**) Corresponding typical cross-section analysis of the **GC3** aggregates on Au (111). (**c**) 300 x 300 nm^2 topographic AFM image and (**d**) corresponding adhesion image. (**e**) 150 x 150 nm^2 topographic AFM image. (**f**) Isometric 2D power spectral density analysis performed on (**e**).

These results were consistent with the presence of a porous in-plane monolayer, where the inner pore diameter was measured as 1.93 nm and 2.31 nm represented the length of the assembled molecule, which also delimited the outer pore diameter. These

experiments were proof of the formation of a porous layer (even if there were regions of bilayers and multilayers) whose dimensions were consistent with those expected for our molecules assembled as cyclic tetramers.

In a next experiment, the surface pressure was decreased in the transference process to see if the formation of single monolayers could be favored (Figure 107).

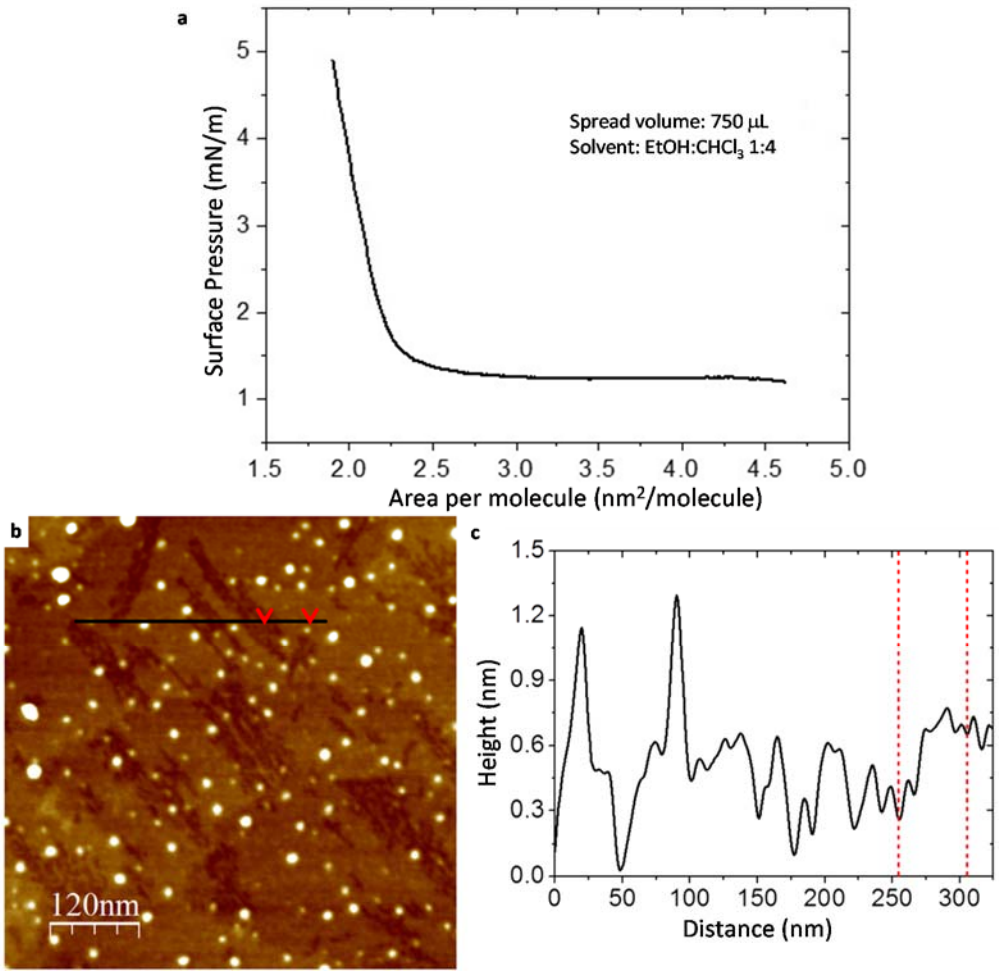


Figure 107. Characterization of a LB film transferred onto Au (111) at 5 mN/m from a monolayer prepared by spreading 750 µL of a GC3 4 µM solution in EtOH:$CHCl_3$ (1:4). (**a**) Surface pressure isotherm. (**b**) Topographic AFM image. (**c**) Corresponding typical cross-section analysis of the **GC3** aggregates on Au (111).

In these experimental conditions, no flakes were detected over the surface. The presence of an incomplete monolayer (0.33 nm high, red arrows) with narrow and elongated fractures which left free zones of the substrate, could be clearly observed in the AFM images. This layer was also decorated with some nanoparticles ranging 0.5–1.5 nm

high and 8–15 nm wide, probably already present in the solution. This LB film was also studied by STM (Figure 108).

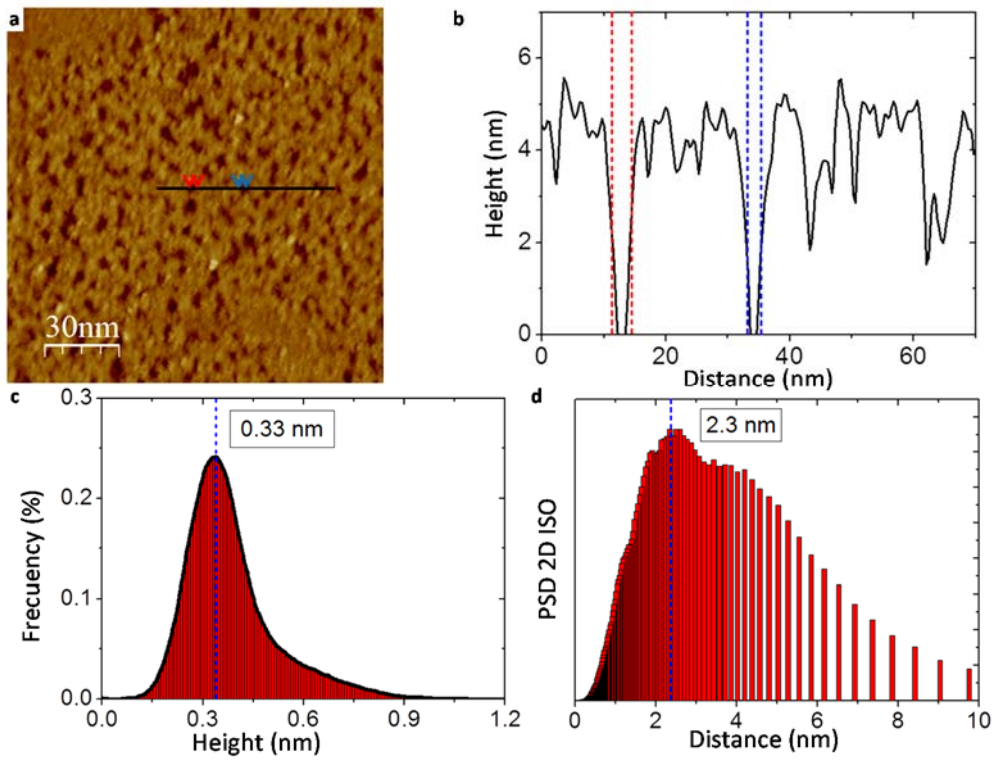


Figure 108. Characterization of a LB film transferred onto Au (111) at 5 mN/m from a monolayer prepared by spreading 750 µL of a GC3 4 µM solution in EtOH:$CHCl_3$ (1:4). (**a**) 150 x 150 nm^2 STM image. (**b**) Corresponding typical cross-section analysis. (**c**) Histogram showing the height value distribution. (**d**) Isometric 2D power spectral density analysis performed on (**a**).

The STM images of the adsorbed layer on a chosen area with no nanoparticles showed an ultrathin porous-shaped organic monolayer whose average height was 0.33 nm, similar to the layer-layer separation on graphite, which is characteristic of aromatic moieties placed parallel to the Au (111) surface. The power spectral density analysis exhibited a maximum at 2.3 nm and also a significant major shoulder at 1.9 nm. These results were, once again, consistent with the presence of a porous in-plane monolayer with a measured inner pore diameter of 1.9 nm and outer pore diameter of 2.3 nm.

4.3. Summary and Conclusions.

The experimental conditions for the preparation of Langmuir films at the water-air interface were optimized for monomer **GC3**. **GC3** molecules exhibited a large tendency to aggregate in chloroform and freshly prepared solutions containing 20% ethanol were required to obtain reproducible results. EtOH was proven to be an adequate co-solvent, miscible with water, prevented strong aggregation and with a sufficient dispersion coefficient. In fact, the aggregation of the molecules, presumably already present in solution, was responsible for the nanoparticles observed later on in the AFM images. These initial aggregates were difficult to totally break and further optimization of the experimental conditions will be carried out so as to obtain clearer solutions for the preparation of LB films. The surface pressure of transference was also proven to be a very important parameter to correct, in order to obtain appropriate LB films for their characterization by AFM and STM techniques. Low surface pressures of transference seem to favor monolayers and avoid the presence of multilayered films. Indeed, high surface pressures of transference were found responsible for the observation of thick films in the AFM images where the porosity of the material could not be resolved.

All these preliminary results seem to indicate that we were able to reproduce at the water-air interface and by vertical deposition, the surface organization obtained at the 1:1 OA:TCB/HOPG interface with **GC3** (see Chapter 2, section **2.2.1.**). In the optimized experimental conditions, monolayers with a porous structure, displaying a measured inner pore diameter of *ca.* 1.9 nm and an external one of *ca.* 2.3 nm were obtained, in agreement with the previous results (see Figure 57, section **2.2.1.**). However, better resolved STM images need to be taken. The first steps toward controlled multilayered materials were given and monolayers, bilayers, as well as three-layered systems could be measured.

Such structures could be identified onto Au (111), which was found to be a better substrate than HOPG and afforded more interesting results. However, further studies need to be carried out onto HOPG. By spreading onto the subphase clearer solutions, with the help of a suitable amount of polar co-colvent, without large particle-like aggregates, and using lower surface pressures of transference, improved results might be achieved. More studies are still required to further determine the optimal conditions for the preparation of these LB films. Additionally, the characterization of Langmuir-Schaefer films (horizontally deposited) will also be explored.

4.4. Experimental Section.

4.4.1. General Methods.

All the solvents used for the preparation of the LB films, as well as the acetone and other solvents used for cleaning purposes were free of surfactants.

4.4.2. LB Films Preparation.

Six Langmuir-Blodgett troughs have been used for the preparation of LB films (Figure 109). Two of them were home-made and presented Teflon™ containers (210 x 460 mm^2) and one mobile barrier each. Surface pressure was determined by a Wilhelmy plate, using a filter paper sheet (19 × 25 mm^2). Both are equipped with a dipping mechanism that holds the substrate and enables controlled deposition cycles. The Wilhelmy plate, the mobile barrier and the dipping mechanism are connected to a computer that monitors the surface pressure values, the area *per* molecule, as well as the programmed transference cycles. The commercial LB troughs are: (1) a NIMA-702BAM, with a 720 × 100 mm^2 container and without dipping mechanism; (2) a KSV-2000, (3) a KSV-NIMA KN 2001 (195 x 50 mm^2) for smaller films prepared out of valuable compounds; and (4) a KSV-NIMA KN 2003 (580 x 145 mm^2). The NIMA-702BAM was dedicated to the specific study of the monolayer formation at the air-liquid interface. The KSV-2000 has two 780 × 120 mm^2 compartments and allows the fabrication of mixed LB films with distinct materials and the alternation of monolayers. All the LB troughs are stored to avoid contamination, kept in a clean laboratory with constant temperature and on top of a vibration free table. The LB troughs are thoroughly cleaned in five steps: (1) the container is first washed with acetone and let to dry (*ca.* 15 min); (2) ethanol is then used and let to evaporate; (3) the container is rinsed in a third step with chloroform and let to dry; (4) three fillying/emptying cycles with ultrapure water (Millipore, Milli-Q, resistivity 18.2 MΩ.cm) are performed; and (5) a blank register of the surface pressure upon compression, without spreading any molecules is, finally, carried out in order to confirm the absence of surfactants or impurities at the water-air interface. The final surface pressure of the process has to be under 1 mN/m.

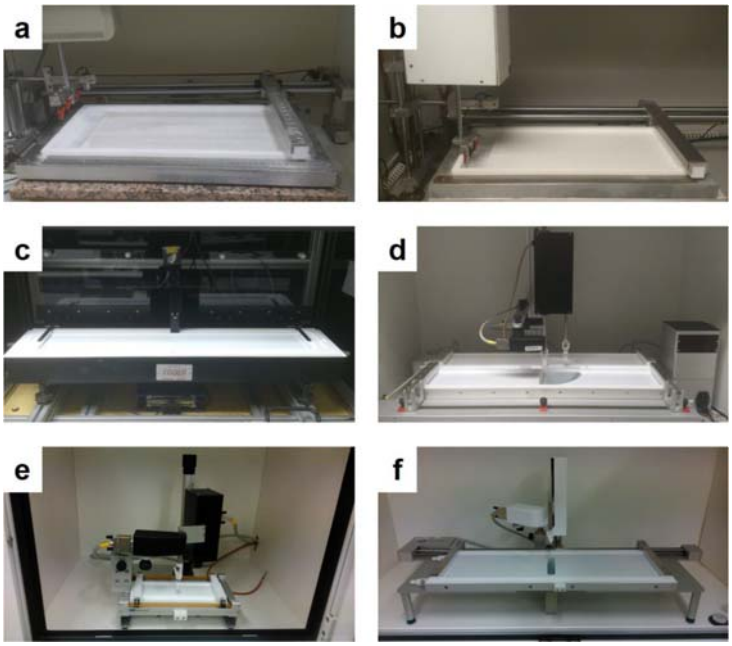


Figure 109. Used Langmuir-Blodgett troughs. (**a**,**b**) Home-made LB troughs; (**c**) NIMA 702BAM; (**d**) KSV 2000; (**e**) KSV NIMA KN 2001 and (**f**) KSV NIMA KN 2003.

4.4.3. LB Films Characterization.

Quantitative mapping was performed at room temperature with a Multimode 8 (Veeco) and using a Nanoscope V controller from Bruker (Figure 110). Samples were imaged by using an Atomic Force Microscope operating in PF-QNM in air at a scan rate of 1.0–1.2 Hz. The AFM probes used in these studies were ScanAsyst-Air-HR (Veeco), with a resonant frequency of 130 kHz, spring constant of 0.4 N/m, and nominal radius of 10 nm. The loading forces during the measurements ranged 300–600 pN. These AFM tips (Figure 110c) were calibrated initially in order to set their deflection sensitivity, *k*; and tip radius.

Structural data and electronic properties of the **GC3** films were also determined by STM and STS, respectively, using a Multimode 8 (Veeco) equipped with a Nanoscope V controller operating in air at room temperature. Pt/Ir tips were used for these measurements.

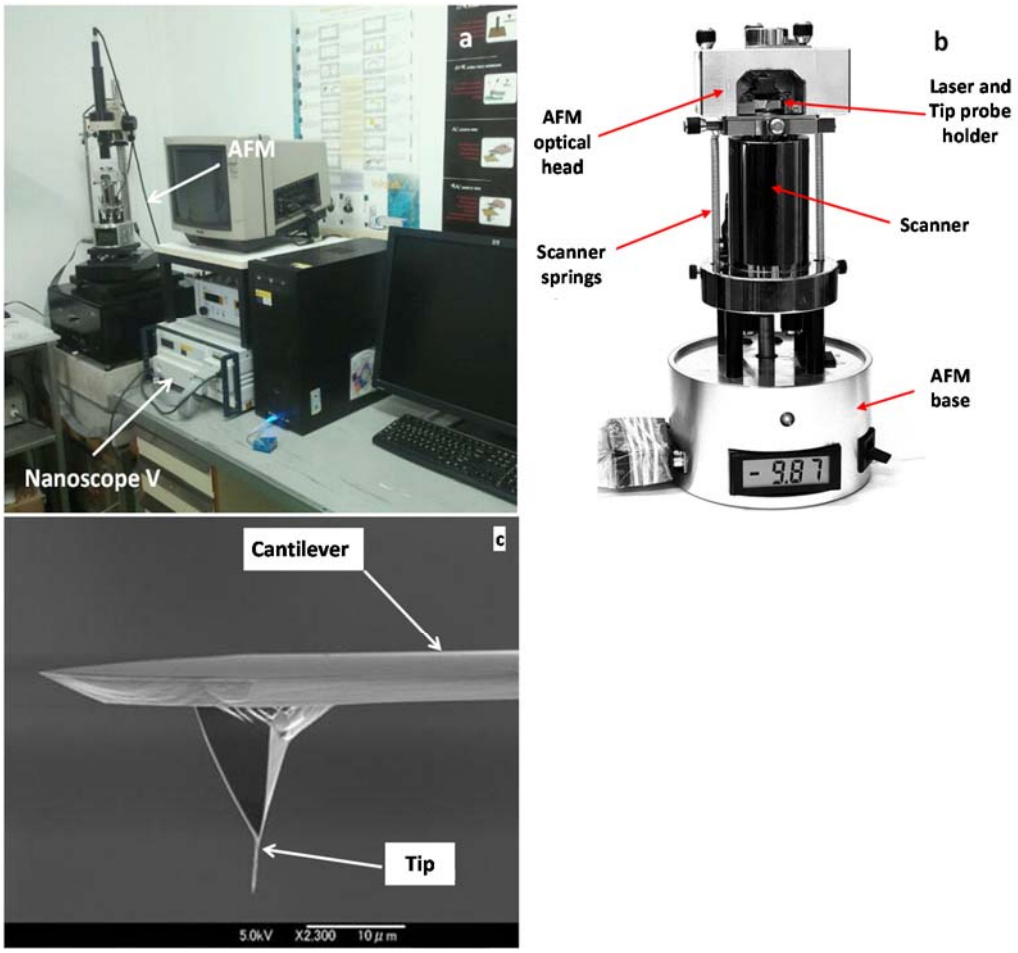


Figure 110. Photograph of the AFM equipment that was used for the characterization of the LB films.

4.4.4. Synthesis and Characterization.

The synthesis and characterization of monomer **GC3** was described in Chapter 2 (section **2.4.2.**).

Chapter 5.

Toward H-bonded Organic Frameworks: Design and Synthesis of Fused Building Blocks.

5.1. Introduction. H-bonded Organic Frameworks.

The synthesis and property investigations of porous materials have become one of the hottest research fields in Materials Science.[267] Their outstanding performance and broad applications, such as gas storage, gas separation, superhydrophobic interfaces, catalysis, energy conversion, energy storage, and optoelectronics, make them tremendously interesting materials. Zeolites[268] and metal–organic frameworks[269] (MOFs) occupy the most privileged spot among them. MOFs are constructed from the combination of metal-containing clusters bound to polytopic linkers.[270] The functional diversity of the organic linker molecules, together with the coordination variety of the metal ion nodes, give access to an enormously large family of MOFs with a wide array of functionalities. On the other hand, porous organic crystalline materials (POCMs),[271] made of only organic compounds, possess lower densities than zeolites and MOFs, because they are composed of light-weight elements. Ordered crystalline POCMs include covalent organic frameworks (COFs)[272] and supramolecular organic frameworks (SOFs).[273] All these polymeric organic networks with discrete pores use the concept of reticular chemistry: topologically designed building blocks are used for the construction of these architectures.

---

[267] a) D. Y. Zhao, J. L. Feng, Q. S. Huo, N. Melosh, G. H. Fredrickson, B. F. Chmelka, G. D. Stucky, *Science* **1998**, *279*, 548–552; b) A. Thomas, *Angew. Chem. Int. Ed.* **2010**, *49*, 8328–8344.

[268] a) M. Moliner, F. Rey, A. Corma, *Angew. Chem. Int. Ed.* **2013**, *52*, 13880–13889; b) J. Jiang, J. Yu, A. Corma, *Angew. Chem. Int. Ed.* **2010**, *49*, 3120–3145; c) M. Moliner, C. Martínez, A. Corma, *Angew. Chem. Int. Ed.* **2015**, *54*, 3560–3579.

[269] a) H. X. Deng, S. Grunder, K. E. Cordova, C. Valente, H. Furukawa, M. Hmadeh, F. Gandara, A. C. Whalley, Z. Liu, S. Asahina, H. Kazumori, M. O'Keeffe, O. Terasaki, J. F. Stoddart, O. M. Yaghi, *Science* **2012**, *336*, 1018–1023; b) H. K. Chae, D. Y. Siberio-Perez, J. Kim, Y. Go, M. Eddaoudi, A. J. Matzger, M. O'Keeffe, O. M. Yaghi, *Nature* **2004**, *427*, 523–527; c) S. C. Xiang, Y. B. He, Z. J. Zhang, H. Wu, W. Zhou, R. Krishna, B. L. Chen, *Nat. Commun.* **2012**, *3*, 954. d) H. C. Zhou, J. R. Long, O. M. Yaghi, *Chem. Rev.* **2012**, *112*, 673–674; e) M. P. Suh, H. J. Park, T. K. Prasad, D. W. Lim, *Chem. Rev.* **2012**, *112*, 782–835; f) L. E. Kreno, K. Leong, O. K. Farha, M. Allendorf, R. P. Van Duyne, J. T. Hupp, *Chem. Rev.* **2012**, *112*, 1105-1125; g) Y. Cui, Y. Yue, G. Qian, B. Chen, *Chem. Rev.* **2012**, *112*, 1126–1162.

[270] a) J. R. Long, O. M. Yaghi, *Chem. Soc. Rev.* **2009**, *38*, 1213–1214; b) M. O'Keeffe, *Chem. Soc. Rev.* **2009**, *38*, 1215–1217; c) D. J. Tranchemontagne, J. L. Mendoza-Cortés, M. O'Keeffe, O. M. Yaghi, *Chem. Soc. Rev.* **2009**, *38*, 1257–1283; d) J. J. Perry IV, J. A. Perman, M. J. Zaworotko, *Chem. Soc. Rev.* **2009**, *38*, 1400–1417.

[271] a) J. T. A. Jones, T. Hasell, X. F. Wu, J. Bacsa, K. E. Jelfs, M. Schmidtmann, S. Y. Chong, D. J. Adams, A. Trewin, F. Schiffman, F. Cora, B. Slater, A. Steiner, G. M. Day, A. I. Cooper, *Nature* **2011**, *474*, 367–371; b) P. Brunet, M. Simard, J. D. Wuest, *J. Am. Chem. Soc.* **1997**, *119*, 2737–2738; c) M. Simard, D. Su, J. D. Wuest, *J. Am. Chem. Soc.* **1991**, *113*, 4696–4698; d) R. Dawson, A. I. Cooper, D. J. Adams, *Prog. Polym. Sci.* **2012**, *37*, 530–563.

[272] a) X. Feng, X. Ding, D. Jiang, *Chem. Soc. Rev.* **2012**, *41*, 6010–6022; b) S.-Y. Ding, W. Wang, *Chem. Soc. Rev.* **2013**, *42*, 548–568.

[273] a) P. Sozzani, S. Bracco, A. Comotti, L. Ferretti, R. Simonutti, *Angew. Chem. Int. Ed.* **2005**, *44*, 1816–1820; b) W. Xiao, Y. Peng, S. Ding, S. Deng, D. Xu, N. Zhang, *Supramol. Chem.* **2015**, *27*, 719–723; c) L.-L. Tan, H. Li, Y. Tao, S. X.-A. Zhang, B. Wang, Y.-W. Yang, *Adv. Mater.* **2014**, *26*, 7027–7031; d) S.-Q. Xu, X. Zhang, C.-B. Nie, Z.-F. Pang, X.N. Xu, X. Zhao, *Chem. Commun.* **2015**, *51*, 16417—16420; e) R. S. Patil, A. M. Drachnik, H. Kumari, C. L. Barnes, C. A. Deakyne, J. L. Atwood, *Cryst. Growth Des.* **2015**, *15*, 2781–2786; f) B. Li, C. Bai, S. Zhang, X. Zhao, Y. Li, L. Wang, K. Ding, X. Shu, S. Li, L. Ma, *J. Mater. Chem. A* **2015**, *3*, 23788—23798; g) J. Kong, J. Du, J. Wang, Z. Chen, *J. Appl. Cryst.* **2015**, *48*, 909–912; h) K.-D. Zhang, J. Tian, D. Hanifi, Y. Zhang, A. C.-H. Sue, T.-Y. Zhou, L. Zhang, X. Zhao, Y. Liu, Z.-T. Li, *J. Am. Chem. Soc.* **2013**, *135*, 17913–17918; i) J. Lü, C. Perez-Krap, M. Suyetin, N. H. Alsmail, Y. Yan, S. Yang, W. Lewis, E. Bichoutskaia, C. C. Tang, A. J. Blake, R. Cao, M. Schröder, *J. Am. Chem. Soc.* **2014**, *136*, 12828–12831; j) J. Tian, T.-Y. Zhou, S.-C. Zhang, S. Aloni, M. V. Altoe, S.-H. Xie, H. Wang, D.-W. Zhang, X. Zhao, Y. Liu, Z.-T. Li, *Nat. Commun.* **2014**, *5*, 5574; k) L. Zhang, T.-Y. Zhou, J. Tian, H. Wang, D.-W. Zhang, X. Zhao, Y. Liu, Z.-T. Li, *Polym. Chem.* **2014**, *5*, 4715–4721; l) X. Zhang, C.-B. Nie, T.-Y. Zhou, Q.-Y. Qi, J. Fu, X.-Z. Wang, L. Dai, Y. Chen, X. Zhao, *Polym. Chem.* **2015**, *6*, 1923–1927.

COFs are crystalline porous polymers where the atomically precise integration of organic units to create predesigned skeletons and nanopores built *via* strong, but reversible, covalent bonds is possible. Depending on the building block, COFs can be categorized into either two- or three-dimensional architectures. In 2D COFs, the covalently bound framework is restricted to bidimensional sheets. These sheets then stack and form a layered eclipsed structure with periodically aligned columns. On the contrary, 3D COFs display a building block containing a sp^3 carbon or silane atom that allow the framework to grow in the third dimension. The reversibility of dynamic covalent reactions, diversity of building blocks, and geometry retention are three key factors involved in the reticular design and synthesis of COFs. An interesting approach in the design of COFs toward further stabilized structures consists in including intermolecular H-bonding in the structure.[274]

While COFs are relatively robust, SOFs usually exhibit low stabilities because they are ordered aggregates of two or more organic species held together by weak non-covalent interactions such as H-bonding, halogen-bonding, cation–π, π–π stacking or van der Waals interactions. Among them, H-bonded organic frameworks (HOFs)[275] represent a subcategory where molecules are linked by directional H-bonds and overcomes the problem of low predictability of the crystal structures. Solvent guests usually play an important role in the construction of this supramolecular network and, once they are removed, the HOF system is usually broken and the H-bonded network collapse. This drawback affects their thermal stability and gas sorption properties. Examples of HOFs with permanent porosities are very rare;[276] and among them, a common feature of these stable HOFs is that the host molecules can form strong H-bonds and π–π stacking interactions with each other. In this case, removal of the solvent has a limited influence on the stability of the assembled supramolecular network. He and coworkers[276b] were the first to report a microporous HOF (**HOF-1**) with permanent porosity and an extraordinary capacity for highly selective adsorptive separation of C_2H_2 and C_2H_4 at ambient temperature (Figure 111a-c). Their design was based on a 2,4-diaminotriazine moiety and included unique features such as framework flexibility, easy purification and regeneration, and high thermal stability. Luo

[274] a) S. Kandambeth, D. B. Shinde, M. K. Panda, B. Lukose, T. Heine, R. Banerjee, *Angew. Chem. Int. Ed.* **2013**, *52*, 13052–13056; b) G. Das, D. B. Shinde, S. Kandambeth, B. P. Biswal, R. Banerjee, *Chem. Commun.* **2014**, *50*, 12615–12618; c) X. Chen, M. Addicoat, E. Jin, L. Zhai, H. Xu, N. Huang, Z. Guo, L. Liu, S. Irle, D. Jiang, *J. Am. Chem. Soc.* **2015**, *137*, 3241–3247; d) C. A. Zentner, H. W. H. Lai, J. T. Greenfield, R. A. Wiscons, M. Zeller, C. F. Campana, O. Talu, S. A. FitzGerald, J. L. C. Rowsell, *Chem. Commun.* **2015**, *51*, 11642–11645.

[275] a) A. Comotti, S. Bracco, G. Distefano, P. Sozzani, *Chem. Commun.* **2009**, 284–286; b) A. Yamamoto, T. Hamada, I. Hisaki, M. MIyata, N. Tohnai, *Angew. Chem. Int. Ed.* **2013**, *52*, 1709 –1712; c) J. Lü, L.-W. Han, J.-X. LIn, R. Cao, *Cryst. Growth Des.* **2011**, *11*, 3551–3557; d) Y. Mizobe, M. Miyata, I. Hisaki, N. Tohnai, *Chem. Lett.* **2007**, *36*, 280–281; e) V. A. Russell, C. C. Evans, W. Li, M. D. Ward, *Science* **1997**, *276*, 575–579; f) N. Tohnai, Y. Mizobe, M. Doi, S.-i. Sukata, T. Hinoue, T. Yuge, I. Hisaki, Y. Matsukawa, M. Miyata, *Angew. Chem. Int. Ed.* **2007**, *46*, 2220–2223; g) X. Hou, Z. Wang, M. Overby, A. Ugrinov, C. Oian, R. Singh, Q. R. Chu, *Chem. Commun.* **2014**, *50*, 5209—5211.

[276] a) A. I. Cooper, *Angew. Chem. Int. Ed.* **2012**, *51*, 7892–7894; b) Y. B. He, S. C. Xiang, B. L. Chen, *J. Am. Chem. Soc.* **2011**, *133*, 14570–14573; c) W. Yang, A. Greenaway, X. Lin, R. Matsuda, A. J. Blake, C. Wilson, W. Lewis, P. Hubberstey, S. Kitagawa, N. R. Champness, M. Schröder, *J. Am. Chem. Soc.* **2010**, *132*, 14457–14469; d) M. Mastalerz, I. M. Oppel, *Angew. Chem. Int. Ed.* **2012**, *51*, 5252–5255; e) X.-Z. Luo, X.-J. Jia, J.-H. Deng, J.-L. Zhong, H.-J. Liu, K.-J. Wang, D.-C. Zhong, *J. Am. Chem. Soc.* **2013**, *135*, 11684–11687; f) P. Li, Y. He, J. Guang, L. Weng, J. C.-G. Zhao, S. Xiang, B. Chen, *J. Am. Chem. Soc.* **2014**, *136*, 547–549.

et al.[276e] went a step further and prepared a system (**HOF-2**) (Figure 111d-f) not only thermally robust but also stable in water and common organic solvents. Desolvated **HOF-2** exhibited high CO_2 adsorption as well as highly selective CO_2 and benzene adsorption at ambient temperature. Particularly interesting is the work published again by the group of Chen[276f] in 2014, where their 2,4-diaminotriazinyl-based homochiral system **HOF-3** (Figure 111g-i) afforded highly enantioselective separation of small molecules for the first time.

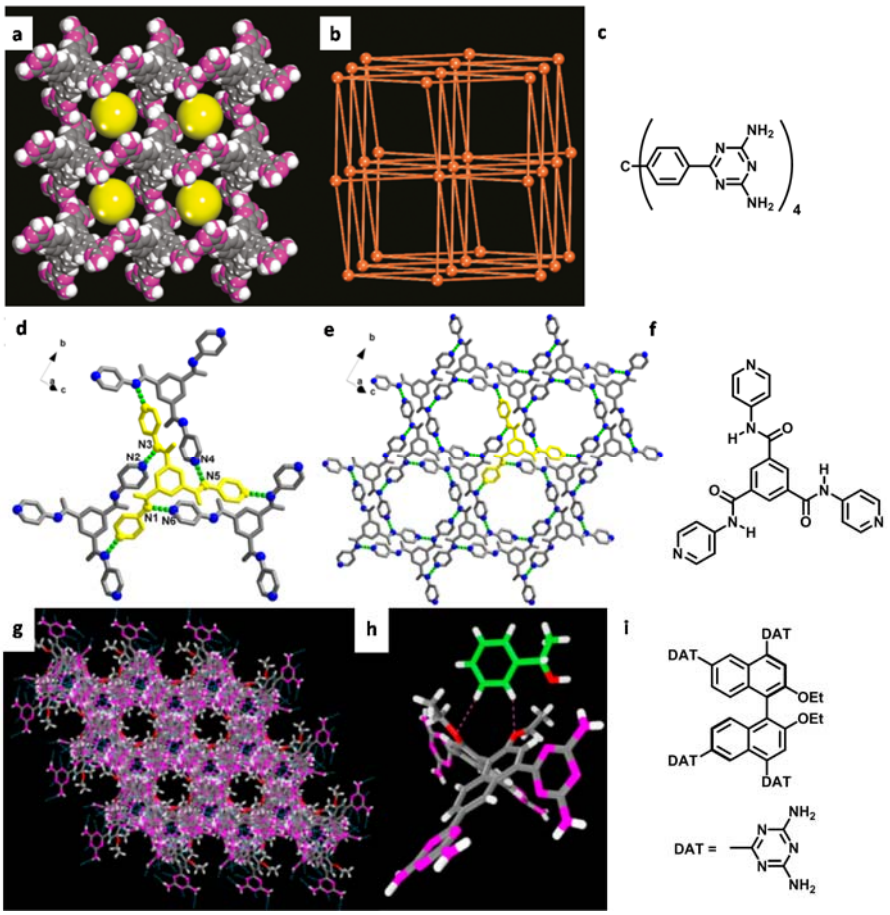


Figure 111. X-ray crystal structures of the systems HOF-1 to HOF-3. (**a**) 1D channels along the *c* axis with a size of ~8.2 Å (yellow spheres) and (**b**) 3D body-centered cubic network topology along with (**c**) organic building block used to construct **HOF-1**. (**d**) H-bonding interactions observed in **HOF-2**. (**e**) 2D layer structure of **HOF-2**, along with (**f**) the corresponding organic building block. (**g**) X-ray crystal structure of **HOF-3** featuring multiple H-bonding (light-blue dashed lines) among adjacent units to form 3D HOF exhibiting 1D hexagonal pores with a ~4.8 Å diameter along the *c* axis. (**i**) X-ray crystal structure of the chiral cavity of the framework for the specific recognition of (*R*)-1-phenylethanol (C, gray; H, white; N, pink; O, red); along with (**i**) the corresponding organic building block.

Our contribution: use of Watson-Crick interactions for the assembly of HOFs.

Solution processability, simpler characterization, easy purification, straightforward regeneration and reuse by simple recrystallization and absence of metals (which can be sometimes toxic or very expensive) are some of the advantages that HOFs display over MOFs. Because of these issues, we decided to make a contribution to the field based on our molecular design. In Chapter 2, robust nanoporous networks were stabilized onto HOPG through specific Watson-Crick H-bonding between complementary GC nucleobases (see Chapter 2, section **2.2.1.**). The secondary interactions that drove the self-assembly in 2D were identified and corresponded to H-bonding between peripheral H-bonding donor and acceptor groups that did not participate in Watson-Crick bonding. Such secondary interactions are indeed much weaker and less specific than nucleobase-pairing. In this Chapter, we aim at directing self-assembly by covalently fusing cyclic tetramers, so self-assembly in 2D would only rely on the stronger Watson-Crick interactions. These, in principle, more robust 2D networks would be studied onto surfaces (Figure 112, top), as we have done with the simpler monomers. Additionally, this design could favor the stacking of sheets of porous tetramer assemblies into crystalline materials (Figure 112). Chapter 5 is divided in two main parts. The first one addresses the design of different nucleobase fused-monomers for the creation of HOFs; and in the second one, the preliminary synthetic results toward some selected systems will be discussed.

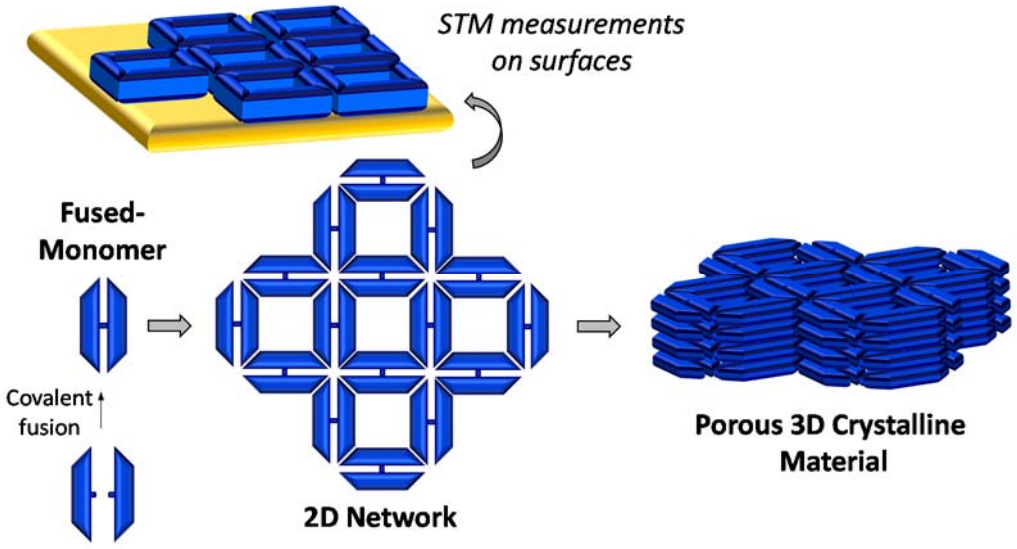


Figure 112. Schematic representation of the objectives of Chapter 5.

5.2. Results and Discussion. Designing HOFs.

The self-assembly strategy followed here for the construction of HOF systems is analogous to the one used in the previous chapters for building nanoporous 2D networks based on discrete macrocycles. In this project, the molecular building blocks that we name "***fused-monomers***" would interact, solely *via* Watson-Crick H-bonding, between ***complementary nucleobases*** to afford 2D porous networks, which would be able to stack and afford 3D materials (related to 2D-COF systems previously described in the literature). The designed units basically comprise two dinucleobase submonomeric units, carrying each a rigid π-conjugated ***central block***, and linked through a suitable ***fusing motif*** (Figure 113).

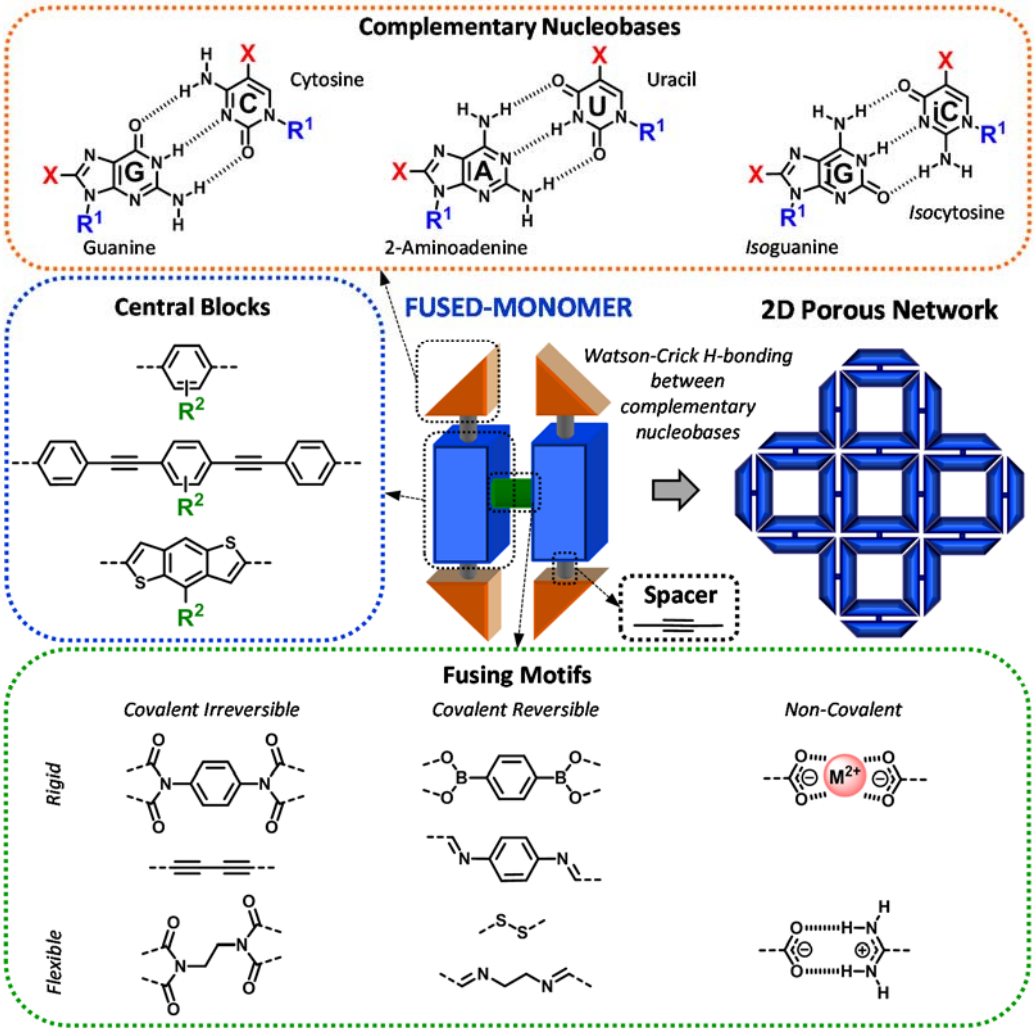


Figure 113. Schematic representation of the structure of the designed fused-monomers.

In this first approximation toward HOFs, submonomeric entities, with an analogous structure to the systems we have previously investigated the most, have been primarily chosen. Therefore, GC ditopic molecules carrying a *p*-diethynyl benzene ring as central block were first selected. On a second level, multiple preliminary designs were considered, taking into account different parameters that concern, not only the primary chemical structure, but the subsequent expected architecture of the final 3D material. The different models were evaluated in accordance to these criteria and following this progressive order:

(a) The **fusing motif**, meaning the chemical functional group that will bind the two dinucleobase derivatives. We evaluated if the considered fusing motifs were (1) reversible or irreversible and (2) conformationally rigid or flexible. We also looked over (3) their length in the different conformations, which influences the final distance between submonomeric units; (4) the ease of synthesis (incorporation in the central blocks and fusing reactions); and (5) their compatibility with the functional groups present in the dinucleobase entities.

(b) The **fused-monomer**. The G-C relative arrangement, leading to two planar conformations: same side or opposite side, was considered. Both conformers were tentatively taken into account and the possible steric interactions evaluated. Then, the most probable structure was selected for the construction of the molecular model.

(c) The sketched **HOFs**. Here, we have estimated the primary pore size (defined by the cyclic tetramer structure) and the secondary pore size (defined by the fusing motif, depending on its nature and the conformation adopted). Let's remark that since all the considered fused-monomers share a common structure based on the G-C pair and the *p*-diethynyl benzene central block, all the systems display the same primary pore size (2.4 nm^2). It is worth mentioning at this point that we seek porous 3D materials where the porosity comes exclusively or mainly from the tailored cavity of the self-assembled cyclic tetramers. Therefore, we will pursue blocking the secondary pores.

Whatever the chemical structure of the linker, our HOF systems will be in any case reversible as their formation relies on specific H-bonding interactions between complementary nucleobases. Nonetheless, a second-order reversible bonding can be incorporated in the system by means of the chosen fusing motif. First of all, we wanted to deal with the simplest case scenario where only Watson-Crick forces contribute to the reversibility of the system. Therefore the four irreversible motifs described below (Figure 114) were considered.

The system **HOF-4** is based on phthalimide derivatives, which are the imide derivatives of phthalic anhydrides (Figure 114a). They are chemically accessible as they can be easily obtained from an amine and phthalic anhydride. It is a rigid linker that should afford in theory a stable HOF material. Different chemical structure modifications may also be easily incorporated, if needed, at the level of the central phenyl ring. The self-assembled cyclic tetramer leaves a primary inner pore of 2.4 nm^2 that will remain available for the

potential selective adsorption of the material. The resulting secondary cavity in the 2D array formed has an area of 2.0 nm^2.

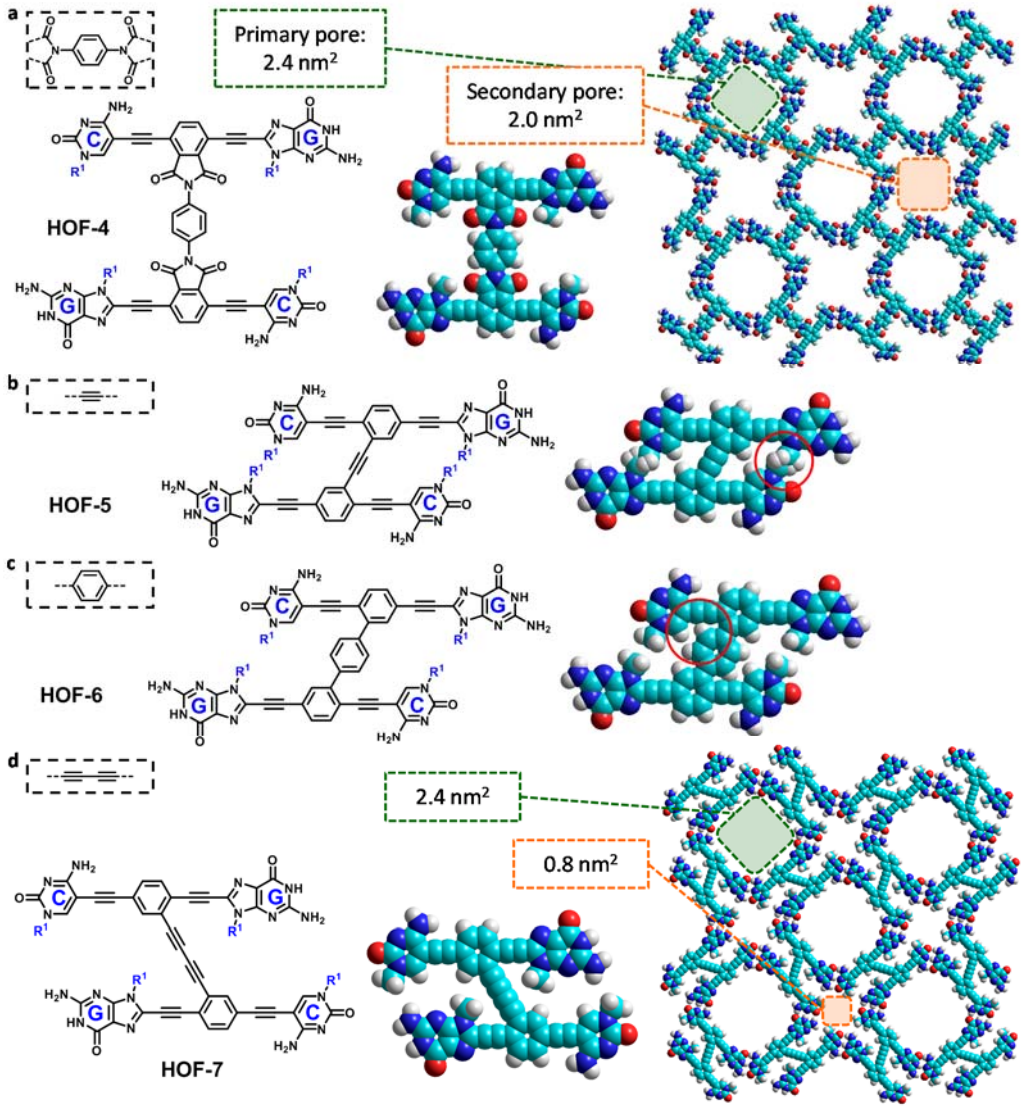


Figure 114. Preliminary considered covalent irreversible target fused-monomers for the generation of HOFs, along with the corresponding primary and secondary pore dimensions. (**a**) **HOF-4** system based on a phthalimide linker, along with the chemical structure of the corresponding fused-monomer. (**b**) Chemical structure of the organic building block considered for **HOF-5** presenting a single triple bond as union and its molecular model showing steric hindrance (red circle) between the alkyl chains. (**c**) **HOF-6** system presenting a *p*-disubstituted phenyl ring as union and its molecular model showing steric hindrance (red circle) in a fully planar conformation. (**d**) Chemical structure of the fused-monomer used for the generation of **HOF-7** based on a diacetylene union.

The fusing motif in **HOF-5** is a rigid ethynyl group that covalently binds two dinucleobase derivatives (Figure 114b). This functional group is easily accessible through Pd-catalyzed reactions, such as Sonogashira or Stille cross-couplings. Two structures were considered: a first one where the triple bond is placed in the position closer to the C derivative; and a second one where the union is closer to the G-nucleobases. Both dispositions resulted in unfavorable steric hindrance, even with short R^1 alkyl chains. Let's precise that the molecular models of the 2D layer were not built when severe steric hindrance was identified, as the formation of the corresponding HOF material would be improbable.

The same two choices were taken into account when placing the phenyl ring linker in the **HOF-6** system (Figure 114c). In a similar manner, both arrangements culminated in an unpromising hindered network and were thus discarded.

Finally, a rigid diacetylene linker was incorporated in the design of the fused-monomer **HOF-7** (Figure 114d). This time, due to the larger distance between submonomeric units, the resulting structure is less sterically hindered than the related **HOF-5**. Again, the linker can be substituted in a neighboring position to the G- or C- derivatives, and this could be defined by synthesis. We have considered the conformation where the fusing motif is placed closer to the purines because it leaves a narrower secondary pore in the 2D network.

Some other designs were targeted where the fusing motif may afford, in addition to the Watson-Crick pairing, a second order of reversibility in the system (Figure 115).

The fused-monomer used for **HOF-8** is based on a reversible boronate ester linker that binds two dinucleobase units (Figure 115a). This molecule was targeted as it is isostructural to **HOF-4** and chemically accessible. Boronate esters have been vastly used for the construction of stable COF materials.[166] Boronic acids can be co-condensed with diol derivatives to afford rigid five-membered boronate ester motifs. Even acetonide-protected catechols can be directly reacted with boronic acid instead of di-alcohols in the presence of the Lewis acid catalyst $BF_3.O$.[177] Again, different desired chemical structure modifications might be incorporated at the level of the central phenyl ring. The dimensions of the self-assembled 2D network remain identical to the ones displayed by **HOF-4**, where the inner cavity has an area of 2.4 nm^2 and the secondary pore measures 2.0 nm^2.

The fusing motif in **HOF-9** is a flexible disulfide bridge that covalently binds two GC submonomeric derivatives (Figure 115b). Different combinations are possible with this system and all of them were considered before selecting the preferred motif. On the one hand, as it was seen in previous examples, the functional group can be placed whether close to G- or C-nucleobases. On the other hand, as this linker is flexible, it can adopt two conformations: a first one where the dinucleobase derivatives are translated with respect to one another, in a parallel direction to the central block; and a second one where the union separate both submonomeric units while maintaining them facing each other, in a

perpendicular direction to the central block. We have represented in Figure 115b the second case scenario in which the linker is fused in a closer position to the purines because it is the only one where no steric hindrance occurred. The resulting densely-packed 2D network displays a secondary pore of 0.6 nm^2.

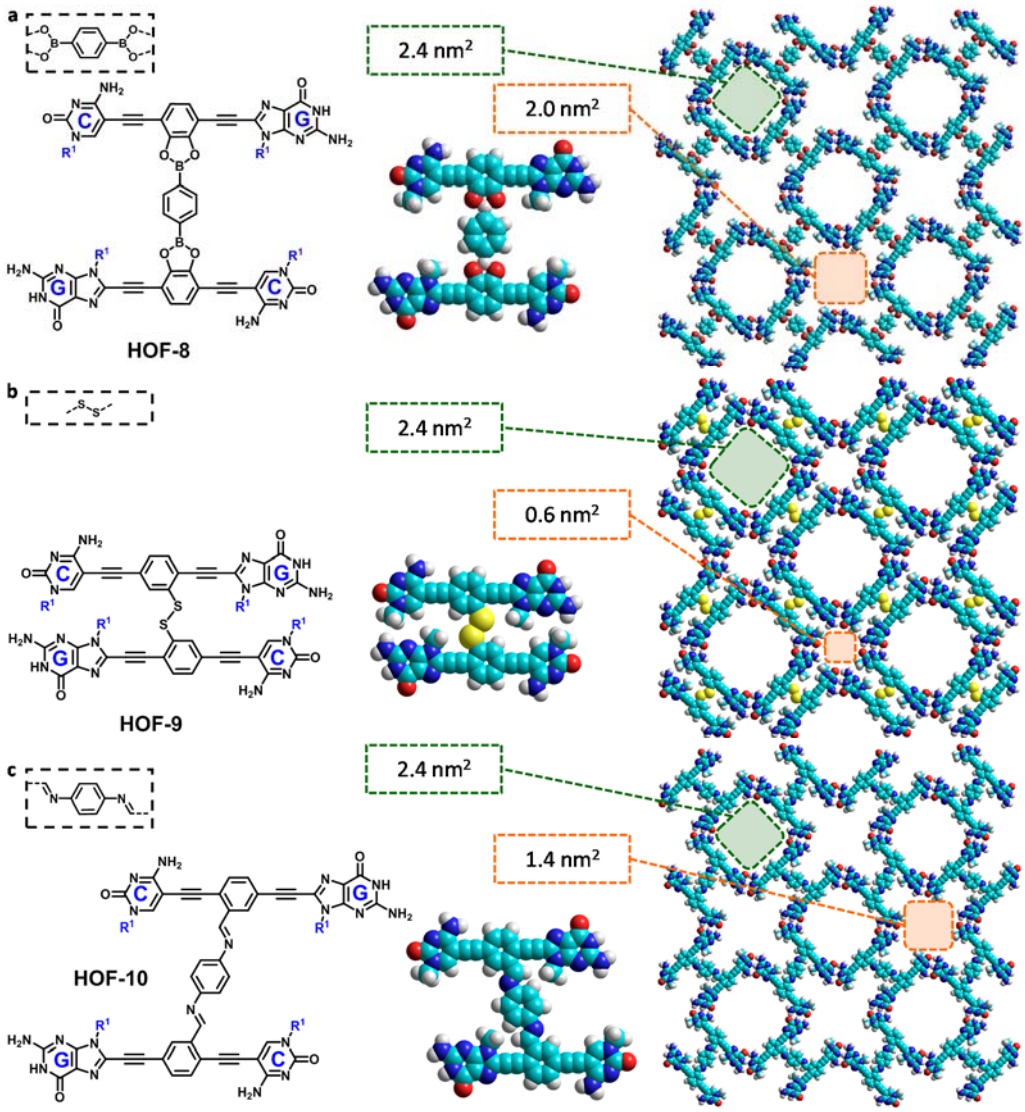


Figure 115. Preliminary considered reversible target fused-monomers for the generation of HOFs, along with the projection of the corresponding 2D layer. (**a**) Fused-monomer used for **HOF-8** based on a boronate ester reversible union. (**b**) Organic building block in **HOF-9** based on a disulfide bridge. (**c**) **HOF-10** system based on a diimine linker.

The final system we studied is based on an imine linker (Figure 115c). Such molecules are accessible through the condensation of aldehyde derivatives with 1,4-diaminobenzene,

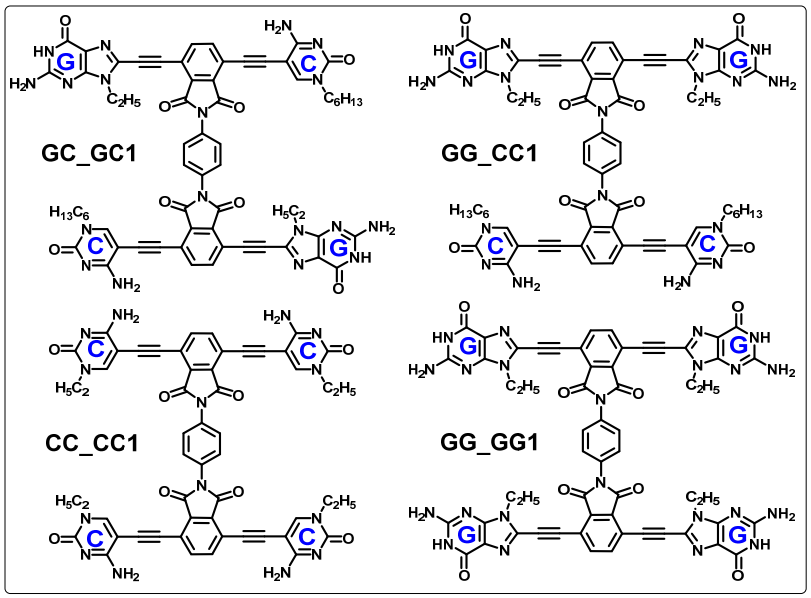
for instance. Other amines can be considered as well, but we chose this *p*-disubstituted phenyl derivative because of its rigidity; and hence the supposed higher stability of the subsequent HOF material. The same four choices that were contemplated for **HOF-9**, arose again when considering **HOF-10**. This time the model outlined a negative steric interaction when locating the fusing motif in a neighboring position to the G-derivatives. Therefore, the fused-monomer where the functionalization of the central block takes place closer to the pyrimidines was selected for building the model in which the secondary pore leaves an area of 1.4 nm^2.

Following the guidelines described above and after simulating the theoretically expected HOF materials, two fusing motifs were preselected. The first one is the non-reversible phthalimide-based fused-monomer **HOF-4** and the second one, its related boronate ester system **HOF-8**. These structures do not present unfavorable steric interactions and are chemically accessible. They do not offer the possibility of additional conformations, like the flexible linkers described in Figure115b,c. Moreover, with these systems, we aim to draw a comparison between isostructural systems only differing in the chemical reversibility of the linker bonds.

We will proceed now discussing the precise design of the selected fused-monomers for the construction of HOF materials.

5.2.1. Phthalimide-based HOFs.

The first task we dealt with in the design of the targeted fused-monomers was the optimization of the alkyl chain lengths. The R^1 tails are useful to us because they block the unspecific secondary pores left between cyclic tetramers, and in this way, a better control on the adsorption properties of the material can be reached. Additionally, they will provide the molecules with sufficient solubility for their adequate manipulation. However, these alkyl chains should not be too long as they would result in unfavorable steric interactions and would prevent the formation of monocrystals. According to our theoretical models, ethyl and hexyl groups resulted well-suited for the functionalization of the G and C derivatives, respectively. Different fused-monomers based on the phthalimide linker were considered, depending on the disposition of the G and C nucleobases (Scheme 28).



Scheme 28. Considered phthalimide-based targets for the generation of HOFs.

One-component HOFs.

Despite the free rotation along the σ-bonds in the axis of the phthalimide linker, G- has been placed in the models opposed to the C- nucleobase in order to alternate the alkyl tails and afford a more sterically favored disposition in **GC_GC1**. However, the molecule has the possibility to rotate around the N–C bonds during assembly.

Molecule **GG_CC1** presents a non-symmetric structure where a GG submonomeric unit is opposed to a CC derivative. Let's note that since the purine derivatives are longer than the pyrimidines, this unsymmetric fused-monomer would afford rectangular-shaped pores in the resulting 2D array.

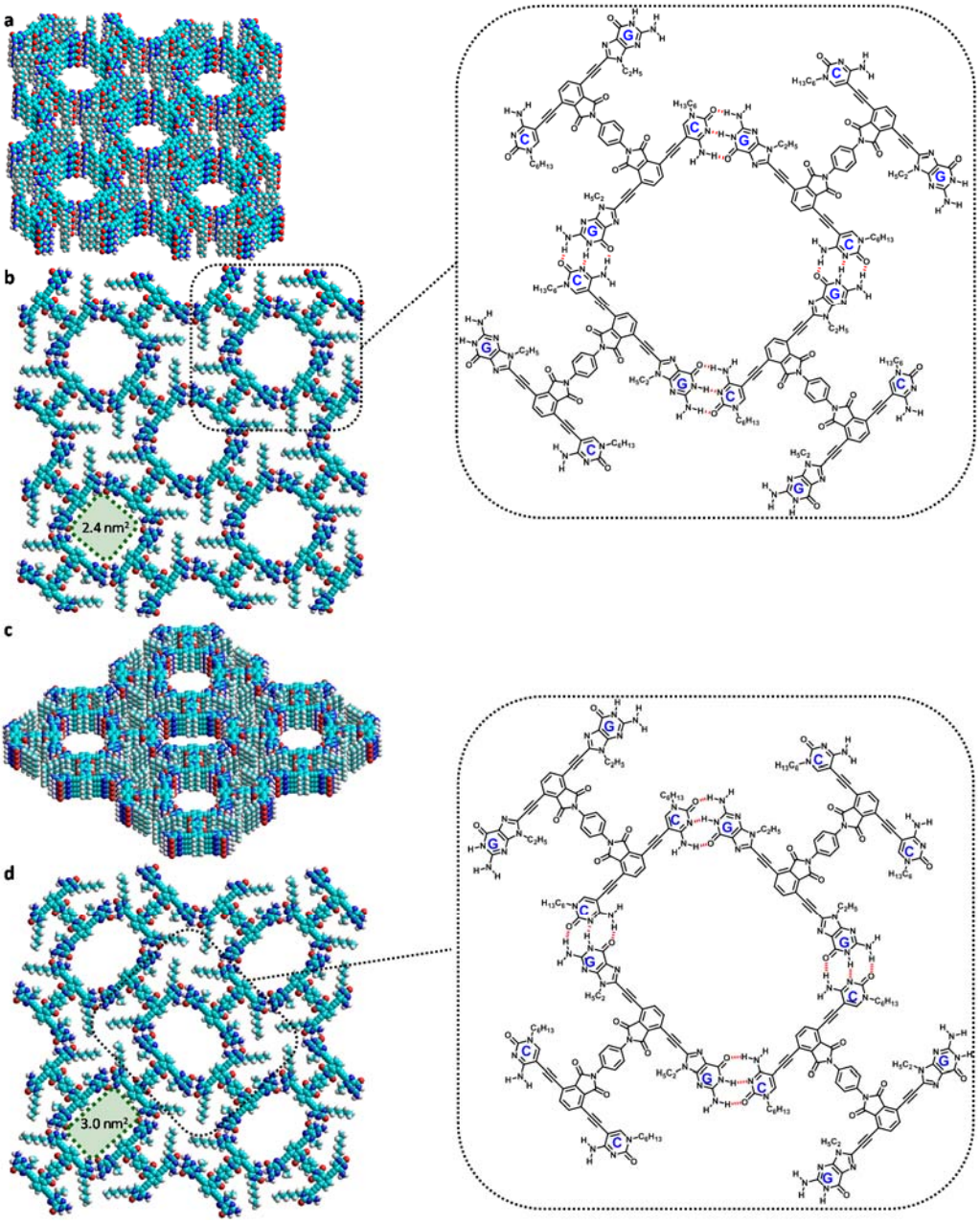


Figure 116. HOF materials formed by GC_GC1 and GG_CC1. (**a**) HOF porous material based on building block **GC_GC1**. (**b**) Corresponding 2D-layer where the secondary cavities have been filled with hexyl chains, along with chemical structure of the repeating unit. (**c**) HOF system based on **GG_CC1**. (**d**) 2D **GG_CC1**-layer showing rectangular-shaped pores, along with chemical structure of the repeating unit.

Other H-bonded structures.

In this first approach we have considered the system in 2D (Figure 116b). Then, we expect successive layers of this material to stack on top of each other, affording a three-dimensional porous HOF (Figure 116a). Nevertheless, a second possibility in which the central block would rotate 90º with respect to the plane of the nucleobases was considered. In this scenario, four dimers can give rise to a discrete rectangular prism (Figure 117a). This species represents the smallest, non-strained system, stabilized by the cooperative action of 24 H-bonds; and hence, should be favored at very low concentrations due to its high *EM* value (see Introduction). However, at higher concentrations, the chelate effect does not apply to that extent, and should lead to the expected organic framework. Additionally, preliminary studies in our research group with prisms based on monomers with a related design but substituted with bulky riboses, showed that these systems rather maintain all the aromatic π-systems (nucleobase and central block) in the same plane.[277] We think that this would also be the case in this scenario and therefore, we expect this cubic species not to be much thermodynamically favored over the 2D layers.

Fused-monomers **CC_CC1** and **GG_GG1** are solely substituted with one type of nucleobase derivative: either cytosine or guanine. These systems can be very interesting for the formation of "chain"-like structures after complementary Watson-Crick H-bonding assembly with a supplementary disubstituted GG- or CC-monomer, respectively. For instance, upon addition of **GG1** to **CC_CC1**, a structure like the one shown in Figure 117b is expected. The complementary nucleobases would recognize each other and associate forming rectangular-shaped cyclic tetramers. Upon concentration, secondary H-bonding interactions between the G-aminopyridine fragments would afford an array of stripes. Let's remark that this 2D cyclic tetramer array is stabilized through covalent interactions in the direction "*x*" of the plane; but *via* supramolecular interactions (*i.e.* H-bonding) in the "*y*" direction. In contrast, the incorporation of monomer $\mathbf{GG_{rib1}}$ with bulkier and solubilizing substituents could avoid this association between stripes and form isolated "tape"-like structures (Figure 117c). Another possibility would be to mix the fused-monomers **CC_CC1** and **GG_GG1**. Upon association, this two-component system would give access to a rectangular-shaped porous HOF (Figure 117e).

[277] Unpublished results.

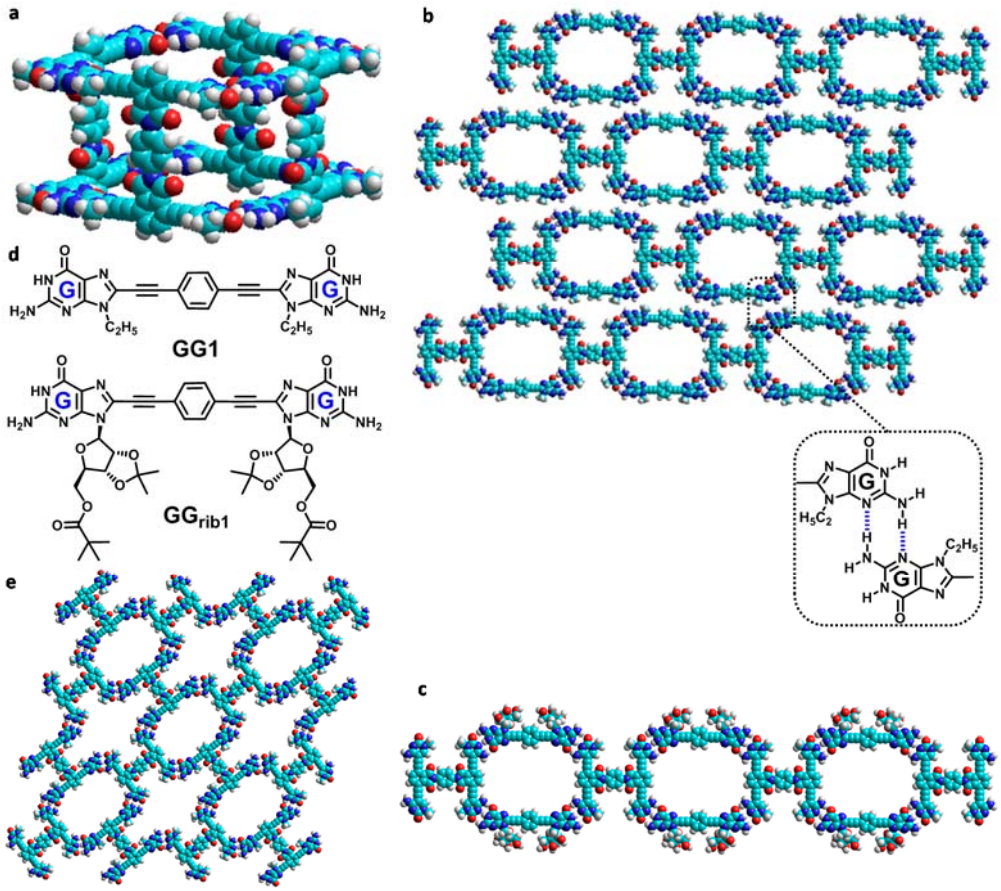


Figure 117. Other structures. (**a**) 3D prism formed by four **GC_GC1** molecules. (**b**) 2D network formed by **CC_CC1** and **GG1**, along with scheme of the secondary stabilizing H-bonding interaction between fibers. (**c**) Single stripe based on the self-assembly of **CC_CC1** and **GG$_{rib1}$**. (**d**) Chemical structures of the considered monomers **GG1** and **GG$_{rib1}$**. (**e**) 2D-layer of the HOF two-component system formed by **GG_GG1** and **CC_CC1**.

Size control.

In this first approach, all the designed fused-monomers were based on a common structure, substituted with a *p*-diethynyl benzene central block. Nonetheless, a wide family of central blocks with different lengths can be incorporated. Since the size of the fused-monomer defines the dimensions of the inner cavity of the porous crystalline material, by tuning the molecular components, a large collection of 3D systems with different adsorption properties (selective toward distinct guest molecules) could be achieved.

Multicomponent systems.

For simplicity reasons, GC systems were considered in a first moment. However, as it was the case in the previous chapters of this Thesis, our approach grants the possibility of mixing different nucleobase derivatives toward multicomponent systems. Here, we could think of the incorporation of both G-C and A-U complementary base-pairs within a fused-monomer for the construction of a HOF material presenting two types of selective cavities with, for instance, different pore size (Figure 118).

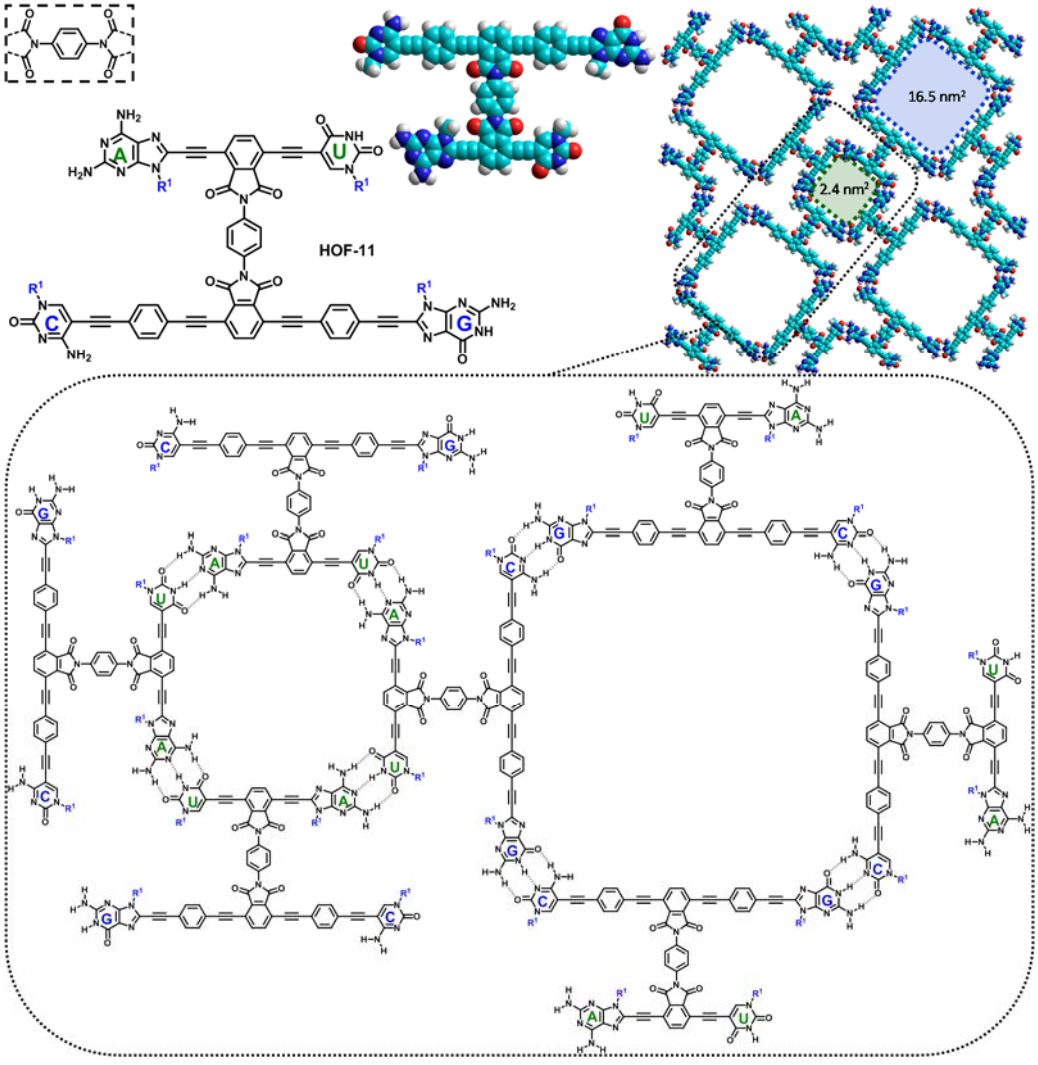
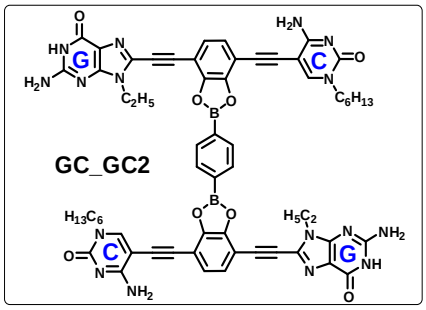


Figure 118. Multicomponent HOF system. Proposed multicomponent fused-monomer and resulting 2D network after self-assembly, corresponding to the section of the attempted HOF material with pores of two different sizes.

5.2.2. Boronate ester-based HOFs.

Fused-monomer **GC_GC2** is based on a boronate-ester linker. This system is isostructural with the phthalimide-based system previously described. Therefore, the same structural considerations were taken into account here. The chosen alkyl tail length for the functionalization of the nucleobase derivatives was identical to **GC_GC1** and $-C_2$ and $-C_6$ chains were adopted (Scheme 29).



Scheme 29. Considered boronate ester-based target **GC_GC2** for the generation of HOFs.

Being the boronate fusing motif reversible in the presence of acid and/or moisture, scrambling phenomena are possible. Only the symmetric **GC_GC2** system was thus attempted this time in order to preliminary evaluate its stability and self-assembly behavior.

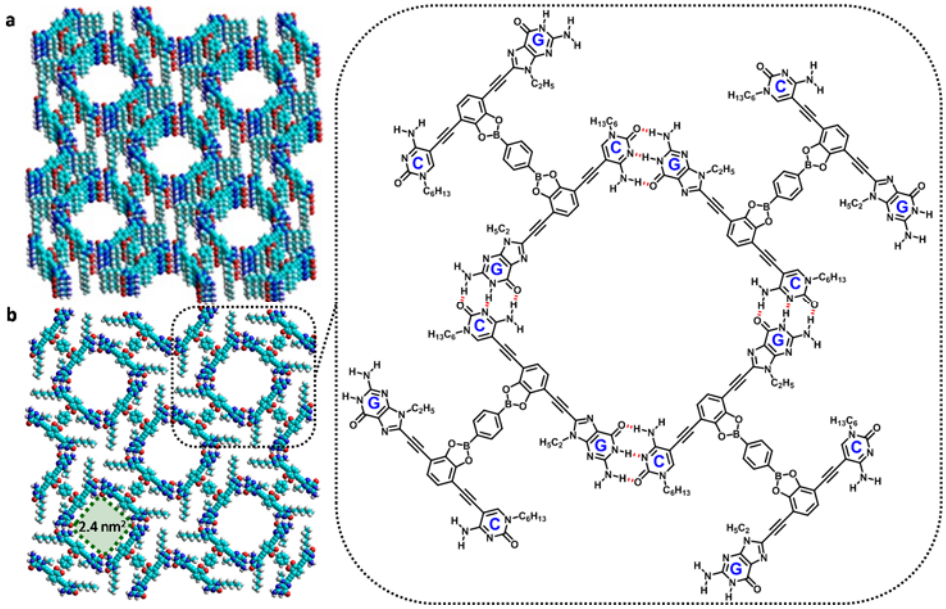


Figure 118. Self-assembly of the fused-monomer GC_GC2. (**a**) 3D HOF material. (**b**) Corresponding 2D-layer, along with chemical structure of the repeating unit formed by four **GC_GC2** molecules.

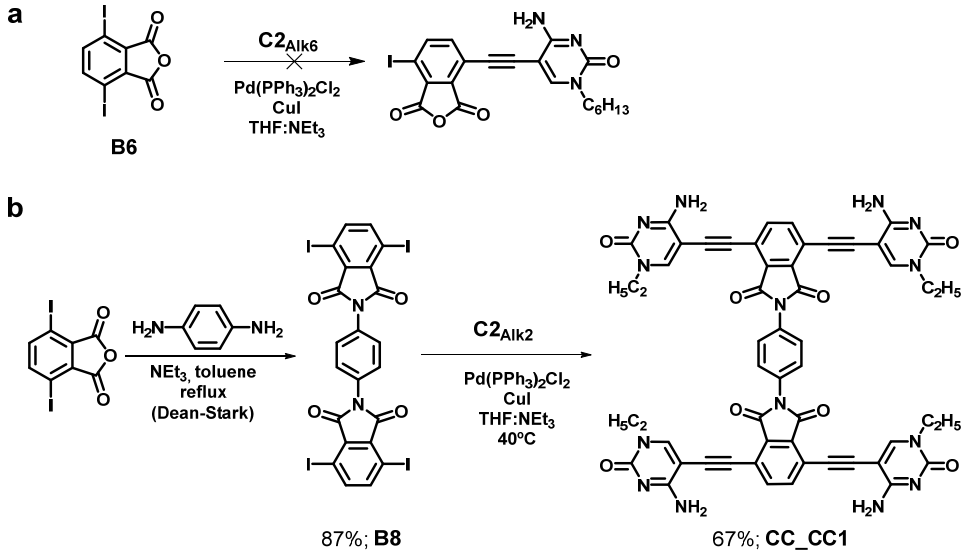
5.3. Results and Discussion. Synthesis of fused-monomers.

We expect the final fused-monomers to be highly insoluble and difficult to manipulate. Therefore, the union between submonomeric entities has been delayed to the last step of the synthetic route. Moreover, this would provide a more convergent route. In order to prepare the desired fused-monomers, a similar synthetic strategy than the one used for single monomers, which are here intermediate species, has been used. This means that the ethynylated nucleobase derivatives will be attached to the selected central blocks *via* Sonogashira coupling reactions. We will focus in this Chapter on the preparation of the different previously introduced fused-monomers. The formation of the HOF systems and their characterization will be considered in the future.

5.3.1. Phthalimide-based HOFs.

In order to prepare the four target molecules with the phthalic anhydride represented in Scheme 28, different synthetic approaches were considered.

Direct Sonogashira coupling. Our first approach consisted in directly subjecting central block **B6** to standard Sonogashira conditions in order to achieve the desired coupled products in a stepwise manner. In that way, the desired nucleobase derivatives could be selectively incorporated to the phthalic acid moiety in a two-step sequence. 1,4-diaminobenzene would be added in the final step to afford the targeted fused-monomers. However, direct cross-coupling conditions with **C2$_{Alk6}$** did not work well on **B6** (Scheme 30a). Due to the sensibility of the phthalic anhydride to the conditions used in column chromatography, the mixture obtained could not be purified (even in neutralized silica gel).

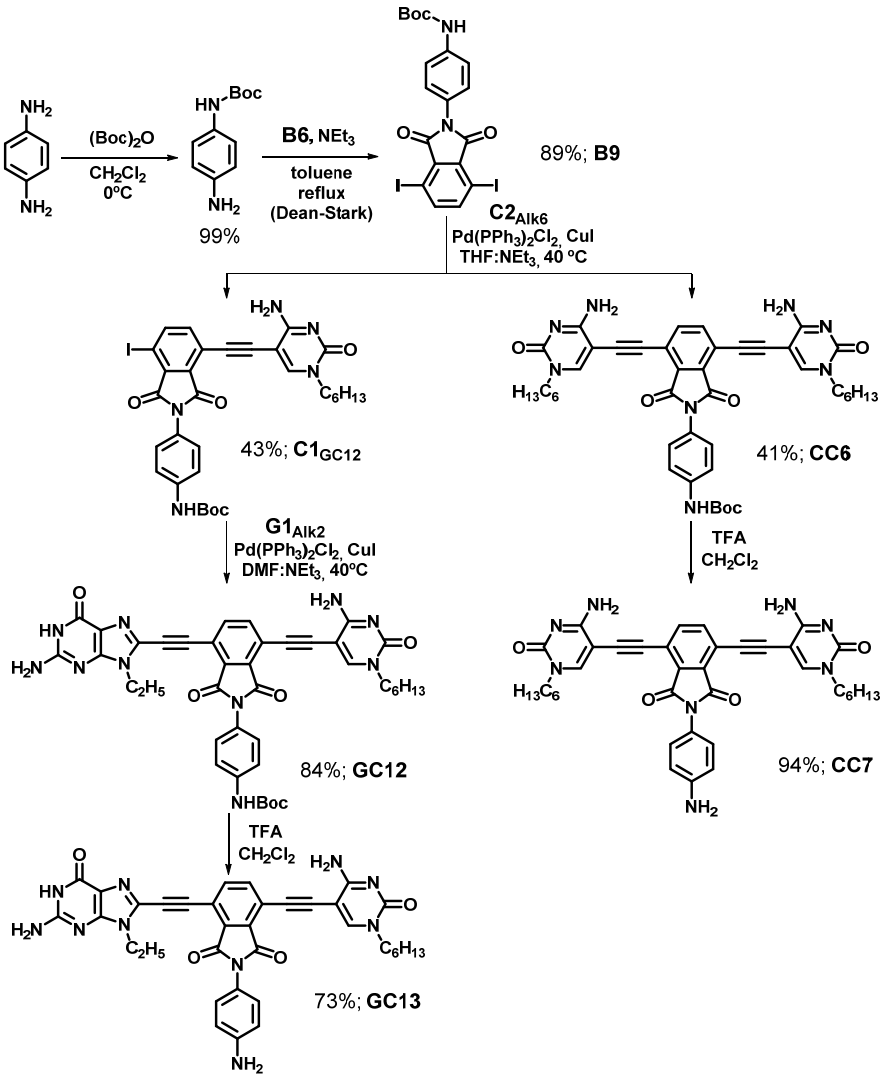


Scheme 30. Synthetic strategies toward phthalimide-based fused-monomers.

Tetraiodinated linker. A second strategy was considered, where central block **B6** was condensed with 1,4-diaminobenzene to afford the tetraiodinated product **B8** with good yields (Scheme 30b). At this point, four consecutive cross-coupling steps were needed to synthesize the desired fused-monomer and none of these reactions held any selectivity. Nonetheless, central block **B8** was mixed with 6 equivalents of $\mathbf{C2_{Alk2}}$, to check its reactivity in Sonogashira coupling reaction conditions. The tetra-substituted **CC_CC1** product was successfully produced in this way. However, when using a lower amount of $\mathbf{C2_{Alk2}}$ equivalents to prepare the mono-substituted derivative, a complex mixture was obtained, containing mono-, di- and tri-coupled products (confirmed by MS analysis) that could not be efficiently purified.

Being quite divergent, in addition to the supplementary purification difficulties, this route was discarded as well.

Stepwise synthesis through the mono-protected diaminobenzene derivative. In a third approach, 1,4-diaminobenzene was protected as a *tert*-butyloxycarbonyl derivative (Scheme 31). This protected aminobenzene derivative was reacted with central block **B6** to afford product **B9** in very good yields. Molecule **B9** was then subjected to two Sonogashira reactions and afforded first the mixture of the mono-coupled product $\mathbf{C1_{GC12}}$ and the doubly-coupled **CC6**, which were purified. Product $\mathbf{G1_{Alk2}}$ was attached in a second coupling reaction with $\mathbf{C1_{GC12}}$ and the Boc protecting group of both the obtained **GC12** and **CC6** monomers was cleaved in the presence of trifluoroacetic acid, yielding **GC13** and **CC7**, respectively.

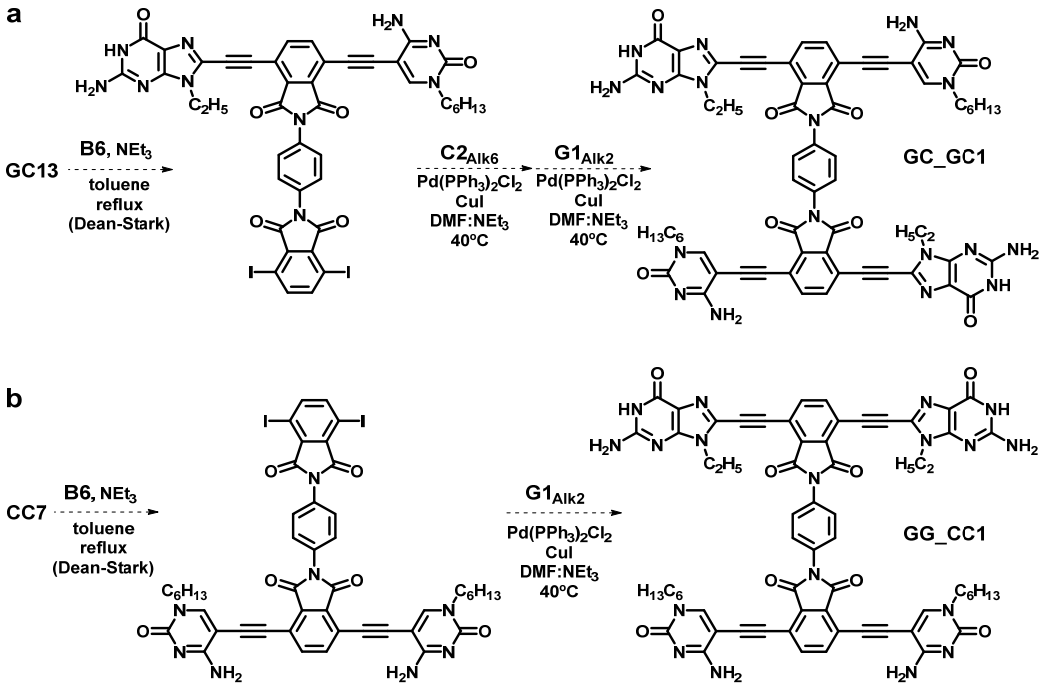


Scheme 31. Synthetic route toward monomers **GC12**, **GC13**, **CC6** and **CC7**.

In order to progress to the final fused-monomers **GC_GC1** and **GG_CC1**, central block **B6** had to be condensed once again to the molecules **GC13** and **CC7** and then successive Sonogashira reactions must be carried out in order to incorporate the desired nucleobase derivatives (Scheme 32).

This approach was good enough for **GG_CC1** but not for **GC_GC1**. It had different disadvantages, such as numerous synthetic steps and it was less efficient in terms of economizing valuable synthetic intermediates. Each Sonogashira reaction step added to the route the production of unwanted subproducts, such as doubly-coupled and homocoupling products (see Chapter 1, section **1.1.3.**) that were not always easy to separate from the

target compounds. Overall, this route was effective but very little convergent. Therefore, a fourth strategy was considered at the same time.



Scheme 32. Proposed synthetic route toward fused-monomers **GC_GC1** and **GG_CC1**.

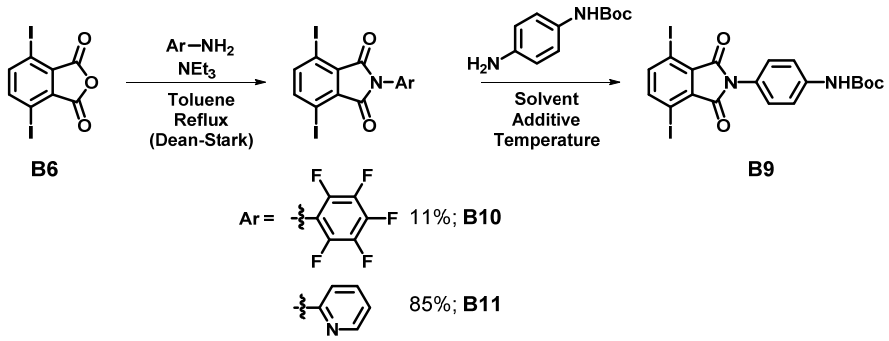
Transimidization. This fourth alternative contemplated the utilization of a transimidization reaction, in which an imide derivative was substituted with another imide group through reaction with an appropriately functionalized amine. This strategy served us for the efficient preparation of fused-monomers in a convergent manner. Indeed, it sufficed to carry out only two consecutive Sonogashira reactions and then perform a last transimidization step.

There are only a few references in the literature that deal with this reaction,[278] even if the transimidization reaction as a method for polyimide synthesis has been known since the 1970s. It was used in the synthesis and handling of ultrahigh pure dianhydrides by blocking the anhydride functionality with an aryl amine.[279] The use of electron deficient aryl amines, such as aminopyridines and aminopyrimidines, allowed for an exchange reaction to take place with an electron-rich (*i.e.* more basic) aryl amine (such as aniline). Bender *et al.*[278a] proposed in 2000 that if the anhydride functionality of a target AB monomer could

[278] a) T. P. Bender, Z. Y. Wang, *J. Polym. Sci. A Polym. Chem.* **2000**, *38*, 3991–3996; b) C. G. Herbert, H. Ghassemi, A. S. Hay, *J. Polym. Sci. A Polym. Chem.* **2000**, *35*, 1095–1104; c) S. T. Hobson, K. J. Shea, Chem. Mater. **1997**, *9*, 616–623; d) J. H. Wengrovius, V. M. Poweel, J. L. Webb, *J. Org. Chem.* **1994**, *59*, 2813–2817.
[279] a) T. Takekoshi, J. E. Kochanowski, *U.S. Patent* 3,850,885, **1974**; b) J. L. Webb, *Eur. Pat.* 132547, **1984**; c) J. L. Webb, *U.S. Patent* 4,578,470, **1986**.

be directly masked by the formation of an imide moiety with an electron deficient heterocyclic amine (*e.g.* 2-aminopyridine or 2-aminopyrimidine), then this would have the effect of rendering the anhydride functionality inert. The AX-type monomer, X being a masking imide group, could then be activated by either heat or a catalyst or both, and then undergo a transimidization reaction with itself to form a polyimide.

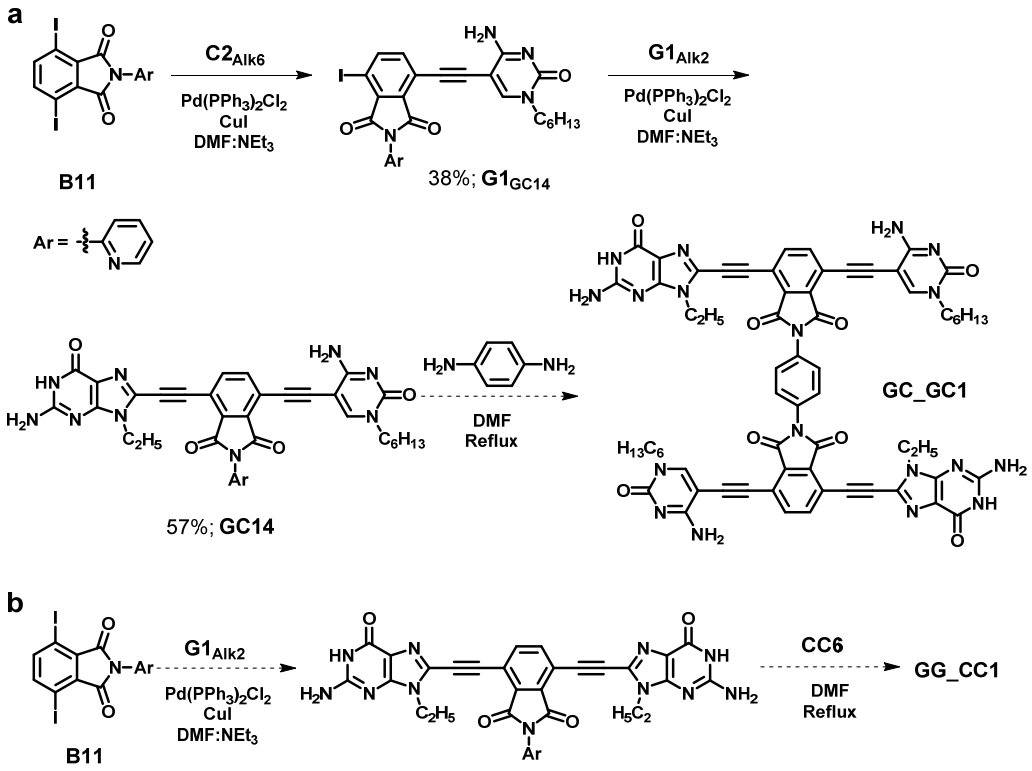
In order to tune up this synthetic strategy, different transimidization conditions were essayed on the iodinated phthalic anhydride derivative **B6** (Scheme 33). Bender *et al.*[278a] worked successfully with 2-aminopyridine to prepare a series of segmented block copolyimides. However, we decided to consider pentafluoroaniline as the first electron deficient amine candidate because it is commercially available, cheap and has a lower boiling point than the 2-aminopyridine reagent (so that the equilibrium can be further shifted by evaporation). The first step of this sequence worked with pentafluoroaniline, affording **XB10** in poor yields. However, the transimidization step proved to be more delicate and numerous conditions were unsuccessfully essayed. The transimidization reaction did not work on the pentafluoro-derivative in xylene under reflux, even when using different catalytic additives such as *p*TsOH. At this point, our efforts were focused on the 2-aminopyridine electron deficient candidate. Again, the reaction with phthalic anhydride afforded the desired product **B11** in good yields, but the second step of the sequence needed further optimization. When using a reflux of cresol (200 ^{o}C), the desired product could be obtained. Yet, cresol has such a high boiling point we decided to essay other solvents and additives. When replacing cresol with propionic acid (141 ^{o}C), the reaction progressed very slowly and the protecting Boc group was cleaved. The reaction could not be improved either by adding catalytic amounts of hydrazine in a reflux of xylene (139 ^{o}C). However, the targeted product could be collected when using *p*TsOH and even without any catalytic reagent. Finally, we were able to obtain the desired product using DMF and we selected this solvent because of its lower boiling point.



Scheme 33. Optimization reaction scheme for the transimidization of phthalimide building blocks.

Once the transimidization conditions were optimized, we addressed the preparation of fused-monomers **GC_GC1** and **GG_CC1**. Our attempted strategy for the preparation of the targeted fused-monomers is shown in the Scheme 34. The 2-aminopyridine substituted

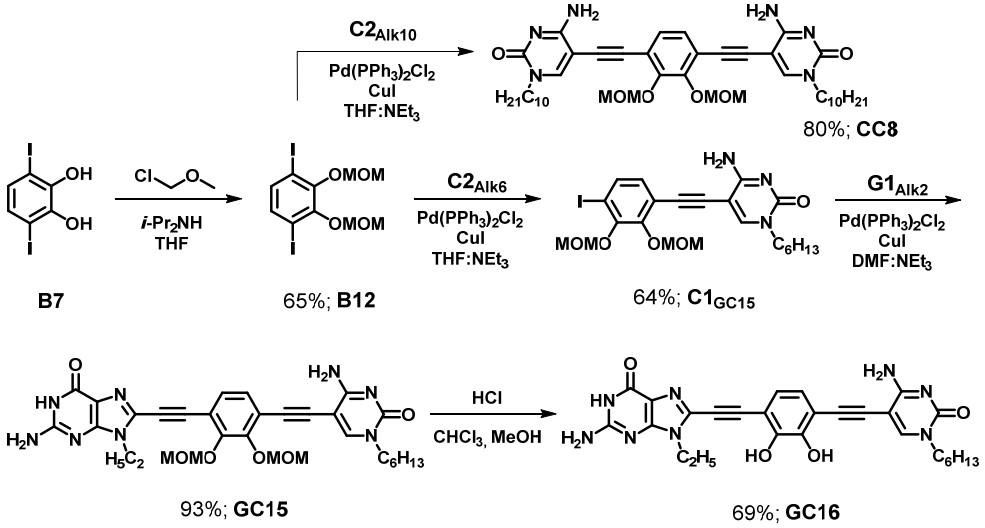
central block **B11** was subjected to the Sonogashira coupling sequence to afford first the monocoupled product **G1$_{GC14}$**, and then, the doubly-coupled **GC14** monomer with moderate yields. Finally, this phthalimide-based disubstituted species functionalized with an electron deficient imide group was subjected to the transimidization reaction, to afford **GC_GC1** with moderate yields. In order to reach symmetric fused-monomers, such as **GC_GC1**, one-step reaction with commercial 1,4-diaminobenzene was sufficient (Scheme 34a). However, unsymmetric fused-monomers like **GG_CC1** or a multicomponent HOF system like the one shown in Figure 118, are only accessible through a transimidization reaction between two dinucleobase submonomeric units, in which one is functionalized with an electron deficient phthalimide derivative and the other one carries the electron richer final desired phtalimide fusing motif (Scheme 34b). It is worth mentioning that molecules **GC13** and **CC7**, synthesized *via* the previous step-wise strategy, will be useful now for the preparation of multicomponent or unsymmetric fused-monomers.



Scheme 34. Synthetic strategy based on a transimidization reaction toward fused-monomers **GC_GC1** and **GG_CC1**.

5.3.2. Boronate ester-based HOFs.

Central block **B7** was subjected to standard Sonogashira reaction conditions with **C2$_{Alk10}$** to check its reactivity. Unfortunately, the coupling did not work on this central block and **B7** was protected as an ether to afford **B12**. This time, the cross-coupling reaction with **C2$_{Alk10}$** was efficient and the protected symmetric monomer **CC8** was obtained in very good yields. The unsymmetric monomer **GC15** was also prepared after two consecutive Sonogashira reactions, first with **C2$_{Alk6}$** and then with **G1$_{Alk2}$**. The cleavage of the methoxymethyl protecting group in the presence of *p*-toluenesulfonic acid did not work. However, monomer **GC16** could be obtained in good yields after subjecting **GC15** to stronger acidic conditions in the presence of HCl (Scheme 35).



Scheme 35. Synthetic route toward monomer **GC16**.

The boronate ester fusing motif being reversible, this system will be first studied in depth with fused-monomer **GC_GC2**. Then, other boronate ester-based systems will be explored. Therefore, monomer **CC8** was left substituted with the methoxymethyl protecting group and will be considered in further experiments. On the other hand, the final monomer **GC16** was isolated in good yields and is ready for condensation with 1,4-phenylenediboronic acid to afford **GC_GC2**. Previous to the generation of the HOF system, monomer **GC16** has been investigated by ^{1}H NMR spectroscopy, which revealed the formation of the corresponding cyclic tetramer. Additional temperature experiments in DMF showed an equilibrium between both species where the cyclic tetrameric species exhibited high stability. After these encouraging preliminary results, fused-monomer **GC_GC2** will be generated in future experiments.

5.4. Conclusions and Perspectives.

Three precursors toward HOF systems have been prepared, based on a rigid and irreversible phthalimide spacer; and a reversible boronic ester linker. Among the first family, four candidates have been considered. The main target is the fused-monomer **GC_GC1** that could afford a robust H-bonded porous organic framework. This system is accessible through the preparation of an unsymmetric GC submonomeric unit. On the other hand, the symmetric **CC7** submonomeric entity is a precursor for **GG_CC1** and would form a rectangular-shaped HOF system. Such a fused-monomer could give access to chain- or tape-like structures upon addition of different GG species. The target molecule **GC_GC2** is isostructural to the fused-monomer **GC_GC1** but relies on a boronate ester fusing motif which is reversible on the appropriate conditions. Notwithstanding, this linker is rigid and we expect **GC_GC2** to form a stable HOF system as well.

The products described in this Chapter, along with their intermediates displayed an extraordinarily insolubility that made syntheses, monitoring of the reactions, as well as the purification protocols, a very difficult task. Nonetheless, we succeeded in the synthesis of a large family of precursors for the preparation of diverse supramolecular 2D and 3D porous systems. The most convergent synthetic routes were always targeted in order to limit the number of reaction steps. This includes the successful incorporation of transimidization reactions within the synthetic route for the development on phthalimide-based HOFs.

The first steps toward a HOF material solely self-assembled through Watson-Crick H-bonding have been given. In this manner, exquisite rational control over the nature and dimensions of the cavities can be reached and therefore, over the potential hosting/adsorption properties of the nanoporous material. On the one hand, these systems will be studied in two-dimensions by different spectroscopic techniques, such as AFM and STM. They can be dropcasted onto different solid substrates or vertically deposited *via* the Langmuir-Blodgett technique. We intend to reach more robust and densely-packed 2D networks than the ones studied in previous chapters, where secondary non-covalent interactions intervened. On the other hand, the formation of 3D materials will be addressed. The number of 2D sheets can be let to grow along the "*z*" axis in a controlled manner for the preparation bi-, tri- to multi-layered systems *via* LB films. Finally, a 3D crystalline porous framework, where Watson-Crick H-bonding is central, will be prepared.

The systems we have described in this Chapter rely on Watson-Crick interactions between complementary nucleobases, and are hence reversible. A second-order reversibility has been incorporated within the **GC_GC2** fused-monomer. The phthalimide linker is irreversible whereas the boronic ester can be in equilibrium with the diol and boronic acid forms under the specific acidic conditions. Dynamic covalent chemistry (DCC) has been extensively used in the construction of COFs. The reversibility of the connections between monomers allows the formation of the most thermodynamically stable structure and facilitates the crystallization of the material, alongside the polymer skeleton forms,

thanks to a self-healing feedback that reduces the incidence of structural defects and assists in the formation of an ordered structure. Kinetically controlled reactions form covalent bonds in an irreversible way and one of the disadvantages of using that approach is that "error-checking" processes do not occur, resulting in linked amorphous organic polymers difficult to crystallize. Therefore, in order to obtain a crystalline and ordered COF or SOF, a reversible reaction must be followed and the molecular building blocks have to be preorganized for the final structure. Inspired by this chemistry, apart from the boronic ester-based **GC_GC2** fused-monomer, **HOF-9** and **HOF-10** systems (see Figure 115b,c) also present an interesting reversible covalent bond formation based on disulfide bridges and imines, respectively. After getting a flavor of the behavior of the prepared systems, we intend the preparation of other HOF systems based on DCC. Additionally, the inclusion of carboxylic acid moieties, for instance, within our monomers for the creation of hybrid hydrogen-bonded and metal-connected organic frameworks based on our molecular design could be envisaged.

Future work will be dedicated to the characterization of these materials, the optimization of the fused-monomer structure and the evaluation of their properties and applications. Further structural and function variations may include controlling cavity size, combining nucleobase pairs (*i.e.* GC and AU), adding internal function to the pores by placing diverse substituents at the central blocks, or evaluating other reversible/irreversible fusing motifs.

5.5. Experimental Section.

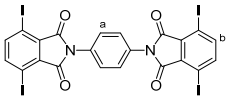
5.4.1. General Methods.

The General Methods detailed in the Experimental Section **1.4.1.** of Chapter 1 are also applicable here. The molecular models were built using Hyperchem™ 8.0 program.

5.4.2. Synthesis and Characterization.

The synthesis and characterization of molecules **C2$_{Alk6}$**, **C2$_{Alk10}$**, **G1$_{Alk2}$**, **B6** and **B7** was described in Chapter 1 (section **1.4.2.**). The *Standard Procedure F* for the Sonogashira coupling with ethynyl-nucleobases described in Chapter 2 (section **2.4.2.**) also applies here.

Phthalic acid derivatives.

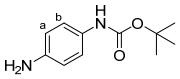


B8. Compound **B6** (1,00 g, 2.62 mmol) was added to a solution of 1,4-diaminobenzene (142 mg, 1.31 mmol) in toluene (75 mL) and NEt_3 (1.65 mL). The mixture was refluxed at 120 °C for 12 h using a Dean-Stark apparatus. The solvent was eliminated *in vacuo*; and then, water (35 mL) and HCl conc. (0.5 mL) were added. The mixture was stirred for 30 min and the resulting solid was filtered and washed with acetonitrile, affording product **B8** as a beige solid (0.99 g, 87%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) =7.96 (s, 4H, Ar (b)), 7.60 (s, 4H, Ar (a)).

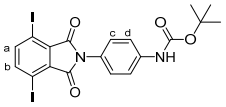
^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.2, 145.7, 133.7, 127.9, 90.4.

HRMS (MALDI, DCTB + NaI): Calculated for $C_{22}H_8I_4N_2Na_1O_4$ $[M+Na]^+$: 894.6555. Found: 894.6562.



***tert*-butyl (4-aminophenyl)carbamate**.[280] A solution of di-*tert*-butyldicarbonate (978 mg, 4.24 mmol) in CH_2Cl_2 (20 mL) was added dropwise at 0 °C to a solution of *p*-phenilenediamine (2.34 g, 21.27 mmol) in CH_2Cl_2 (100 mL). The solution was stirred under argon for 3 h. The solvent was evaporated *in vacuo* and the crude solid was purified by column chromatography on silica gel eluted with cyclohexane/AcOEt (2:1), affording *tert*-butyl(4-aminophenyl)carbamate as a white solid (878 mg, 99%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.77 (bs, 1H, N*H*), 7.06 (d, *J* = 8.2 Hz, 2H, H^b), 6.46 (d, *J* = 7.9 Hz, 2H, H^a), 4.72 (bs, 2H, NH_2), 1.44 (s, 9H, *t*Bu).



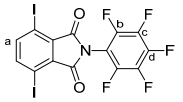
B9. Central block **B6** (202 mg, 505 mmol) was added to a solution of *tert*-butyl (4-aminophenyl)carbamate (115.0 mg, 0.552 mmol) in toluene (20 mL) and NEt_3 (0.45 mL). The mixture was heated at 120 °C in a Dean-Stark apparatus for 5 h. The solvent was eliminated *in vacuo* and the crude solid was purified by column chromatography on silica gel eluted with cyclohexane/ethyl acetate (10:1); affording **B9** as a white solid (265.8 mg, 89%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 9.56 (bs, 1H, N*H*), 7.93 (s, 2H, H^a, H^b), 7.57 (d, *J* = 8.7 Hz, 2H, , H^d), 7.30 (d, *J* = 8.2 Hz, 2H, H^c), 1.50 (s, 9H, *t*Bu).

^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 164.4, 152.7, 145.6, 139.5, 133.6, 127.9, 125.5, 118.1, 90.2, 79.3, 28.1.

[280] M. Togashi, Y. Urano, H. Kojima, T. Terai, K. Hanaoka, K. Igarashi, Y. Hirata, T. Nagano, *Org. Lett.* **2010**, *12*, 1704–1707.

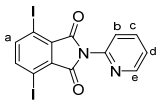
HRMS (MALDI, DCTB + NaI): Calculated for $C_{19}H_{16}I_2N_2NaO_4$ [M+Na]$^+$: 612.9092. Found: 612.9069.



B10. Central block **B6** (100 mg, 0.25 mmol) was added to a solution of pentafluoroaniline (46.7 mg, 0.26 mmol) in toluene (15 mL) and NEt_3 (0.5 mL). The mixture was heated at 120°C in a Dean-Stark apparatus for 24 h. The solvent was eliminated *in vacuo* and the crude solid was purified by column chromatography on silica gel eluted with cyclohexane/AcOEt (10:1), affording **B10** as a white solid (15.4 mg, 11%).

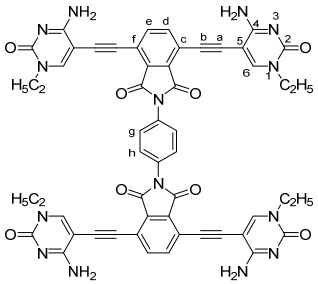
^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.88 (s, 2H, H^a).

^{19}F NMR (300 MHz, $CDCl_3$) δ(ppm) = -141.97 (m, 2F, F^b), -150.29 (m, 1F, F^d), -160.58 (m, 2F, F^c).



B11. Central block **B6** (784 mg, 1.96 mmol) was added to a solution of 2-aminopyridine (200 mg, 2.13 mmol) in toluene (30 mL) and NEt_3 (0.45 mL). The mixture was heated at 120°C in a Dean-Stark apparatus for 24 h. The solvent was eliminated *in vacuo* and the crude solid was purified by column chromatography on silica gel eluted with $CHCl_3$, affording **B11** as a beige solid (790 mg, 85%).

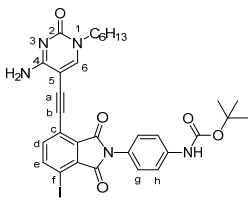
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.65 (ddd, J = 4.8, 2.0, 1.0 Hz, 1H, H^e), 8.06 (td, J = 7.8, 1.9 Hz, 1H, H^c), 7.97 (s, 2H, H^a), 7.60-7.50 (m, 2H, H^b, H^d).



CC_CC1. CC_CC1 was prepared following the *Standard Procedure F*. A solution of **C2$_{Alk2}$** (157.4 mg, 0.964 mmol) in the DMF/NEt_3 mixture (14 mL) was slowly added over 7 h to a suspension of **B8** (200.6 mg, 0.230 mmol), $Pd(PPh_3)_2Cl_2$ (29.0 mg, 0.041 mmol) and CuI (3.2 mg, 0,017 mmol) in DMF/NEt_3 (40 mL) at 60 °C. The resulting mixture was stirred overnight at 60 °C. The crude was concentrated *in vacuo* and the resulting solid was washed with acetonitrile and $CHCl_3$ to afford **CC_CC1** as a brown solid (160.6 mg, 67%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.34 (s, 4H, H^6), 8.00 (s, 4H, H^d, H^e), 7.90 (bs, 4H, C^4NH-*H*), 7.68 (s, 4H, H^g, H^h), 7.11 (bs, 4H, C^4NH-*H*), 3.79 (q, J = 8.1 Hz, 8H, $N^1CH_2CH_3$), 1.21 (t, J = 7.0 Hz, 12H, $N^1CH_2CH_3$).

HRMS (MALDI, dithranol): Calculated for $C_{54}H_{41}N_{14}O_8$ [M+H]$^+$: 1013.3226. Found: 1013.3262.

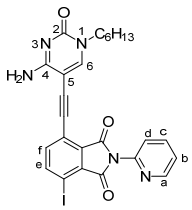


C1$_{GC12}$. **C1$_{GC12}$** was prepared following the *Standard Procedure F* using **C2$_{Alk6}$** (29 mg, 0.13 mmol), **B9** (0.30 g, 0.51 mmol), CuI (0.5 mg, 0.003 mmol) and $Pd(PPh_3)_2Cl_2$ (4 mg, 0.005 mmol) in THF/NEt_3 (20 mL). The reaction was completed overnight. Then, the reaction mixture was concentrated under reduced pressure and washed with cold acetonitrile. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1); affording **C1$_{GC12}$** as a yellow solid (38 mg, 43%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 9.56 (bs, 1H, N*H*Boc), 8.30 (s, 1H, H^6), 8.26 (d, J = 8.2 Hz, 1H, H^e), 7.94 (bs, 1H, C^4NH-*H*), 7.56 (m, 3H, H^d, H^h), 7.32 (d, J = 8.5 Hz, 2H, H^g), 7.10 (bs, 1H, C^4NH-*H*), 3.72 (t, J = 7.3 Hz, 2H, $N^1CH_2C_5H_{11}$), 1.59 (s, 2H, $N^1CH_2CH_2C_4H_9$), 1.50 (s, 9H, *t*Bu), 1.25 (d, J = 8.0 Hz, 6H, $N^1(CH_2)_2(CH_2)_3CH_3$), 0.86 (t, J = 6.5 Hz, 3H, $N^1C_5H_{10}CH_3$).

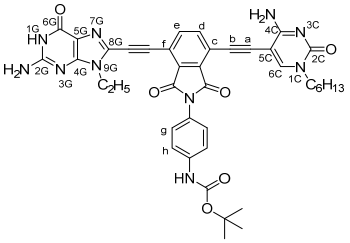
^{13}C NMR (75 MHz, DMSO-d_6) δ(ppm) = 165.6, 165.3, 164.3, 153.8, 152.7, 145.1, 139.7, 132.6, 131.3, 128.4, 128.2, 125.3, 118.5, 118.3, 117.7, 91.2, 90.8, 90.6, 89.6, 79.3, 49.3, 30.8, 28.5, 28.1, 25.5, 21.9, 13.8.

HRMS (MALDI, DCTB): Calculated for $C_{31}H_{33}IN_5O_5$ [M+H]$^+$: 682.1521. Found: 682.1544.



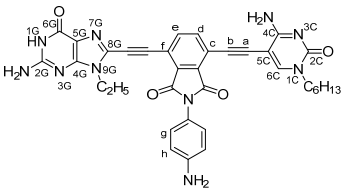
G1$_{GC14}$. G1$_{GC14}$ was prepared following the *Standard Procedure F* using **C2$_{Alk6}$** (77 mg, 0.35 mmol), **B11** (0.5 g, 1.10 mmol), CuI (0.7 mg, 0.004 mmol) and $Pd(PPh_3)_2Cl_2$ (5 mg, 0.008 mmol) in DMF/NEt_3 (60 mL). The reaction was completed overnight. Then, the reaction mixture was concentrated under reduced pressure and washed with cold acetonitrile. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1), affording **G1$_{GC14}$** as a yellow solid (75 mg, 38%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.67 (ddd, J = 4.9, 1.9, 0.8 Hz, 1H, H^a), 8.31 (d, J = 8.2 Hz, 1H, H^f), 8.31 (s, 1H, H^1), 8.08 (td, J = 7.7, 1.8 Hz, 1H, H^c), 7.94 (bs, 1H, C^4NH-H), 7.66–7.51 (m, 3H, H^b, H^d, H^e), 7.02 (bs, 1H, C^4NH-H), 3.73 (t, J = 7.3 Hz, 2H, $N^1CH_2C_5H_{11}$), 1.59 (s, 2H, $N^1CH_2CH_2C_4H_9$), 1.26 (s, 6H, $N^1(CH_2)_2(CH_2)_3CH_3$), 0.85 (t, J = 6.7 Hz, 3H, $N^1C_5H_{10}CH_3$).



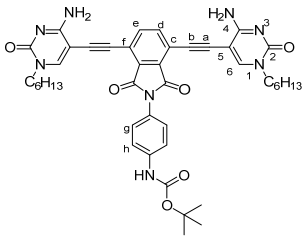
GC12. GC12 was prepared following the *Standard Procedure F* using **G1$_{Alk2}$** (36 mg, 0.18 mmol), **C1$_{GC12}$** (0.12 g, 0.17 mmol), CuI (0.7 mg, 0.003 mmol) and $Pd(PPh_3)_2Cl_2$ (5.4 mg, 0.008 mmol) in DMF/NEt_3 (80 mL). The reaction was stirred under argon at 50 °C and completed overnight. Then, the reaction mixture was concentrated under reduced pressure and washed with cold acetonitrile, affording **GC12** as a brown solid (110 mg, 84%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 10.78 (bs, 1H, NH^{1G}), 9.57 (bs, 1H, NHBoc), 8.35 (s, 1H, H^{6C}), 8.05 (d, J = 8.4 Hz, 1H, H^d), 7.98 (bs, 1H, C^{4C}NH-H), 7.89 (d, J = 7.9 Hz, 1H, H^e), 7.58 (d, J = 8.9 Hz, 2H, H^h), 7.34 (d, J = 8.9 Hz, 2H, H^g), 7.13 (bs, 1H, C^{4C}NH-H), 6.73 (bs, 2H, $C^{2G}NH_2$), 4.29 (q, J = 6.8 Hz, 2H, $N^{9G}CH_2CH_3$), 3.74 (t, J = 7.5 Hz, 2H, $N^{1C}CH_2C_5H_{11}$), 1.60 (m, 2H, $N^{1C}CH_2CH_2C_4H_9$), 1.50 (s, 9H, tBu), 1.32 (t, J = 7.5 Hz, 3H, $N^{9G}CH_2CH_3$), 1.26 (m, 6H, $N^{1C}(CH_2)_2(CH_2)_3CH_3$), 0.86 (t, J = 6.7 Hz, 3H, $N^{1C}C_5H_{10}CH_3$).
HRMS (ESI+): Calculated for $C_{40}H_{43}N_{10}O_6$ $[M+H]^+$: 757.3205. Found: 757.3545.



GC13. TFA (2 mL) was added to a suspension of **GC12** (45.7 mg, 756.5 mmol) in CH_2Cl_2 (10 mL). The mixture was stirred at rt for 29 h. Then, $NaHCO_3$ (sat.) (5 mL) was added and the crude was stirred for 30 min. CH_2Cl_2 was evaporated *in vacuo* and water (5 mL) was added. The suspension was sonicated, filtered and washed with acetonitrile to afford **GC13** as a brown solid (29 mg, 73%).

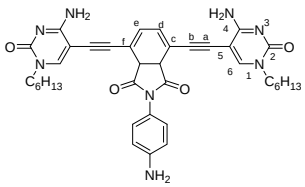
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.78 (bs, 1H, H^{1G}), 8.40 (s, 1H, H^{6C}), 8.21 (bs, 1H, C^{4C}NH-H), 8.05 (d, J = 8.2 Hz, 1H, H^d), 7.89 (d, J = 8.1 Hz, 1H, H^e), 7.42 (bs, 1H, C^{4C}NH-H), 7.12 (d, J = 8.6 Hz, 2H, H^h), 6.76 (d, J = 8.5 Hz, 2H, H^g), 6.72 (bs, 2H, $C^{2G}NH_2$) 4.28 (q, J = 7.3 Hz, 2H, $N^{9G}CH_2CH_3$), 3.75 (t, J = 7.3 Hz, 2H, $N^{1C}CH_2C_5H_{11}$), 1.61 (m, 2H, $N^{1C}CH_2CH_2C_4H_9$), 1.32 (t, J = 7.1 Hz, 3H, $N^{9G}CH_2CH_3$), 1.26 (m, 6H, $N^{1C}(CH_2)_2(CH_2)_3CH_3$), 0.87 (t, 3H, $N^{1C}C_5H_{10}CH_3$).
HRMS (ESI+): Calculated for $C_{35}H_{33}N_{10}O_4$ $[M+H]^+$: 657.2680. Found: 657.2681.



CC6. CC6 was obtained as a second product in the reaction for the preparation of the product **C1$_{GC14}$**. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1), affording **CC6** as a yellow solid (21 mg, 41%).
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 9.54 (bs, 1H, NHBoc), 8.31 (s, 2H, H^6), 7.95 (bs, 2H, NH_2), 7.86 (s, 2H, H^d, H^e), 7.57 (d, J = 8.7 Hz, 2H, H^h), 7.34 (d, J = 8.7 Hz, 2H, H^g), 7.10 (bs, 2H, NH_2), 3.73 (t, J = 7.3 Hz, 4H, $N^1CH_2C_5H_{11}$), 1.64-1.55 (m, 4H, $N^1CH_2CH_2C_4H_9$), 1.50 (s, 9H, tBu), 1.31-1.22 (m, 12H, $N^1(CH_2)_2(CH_2)_3CH_3$), 0.86 (t, J = 6.7 Hz, 6H, $N^1C_5H_{10}CH_3$).

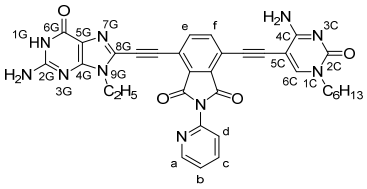
^{13}C NMR (75 MHz, DMSO) δ(ppm) = 166.4, 164.2, 153.7, 152.8. 150.9, 139.8, 130.3, 128.4, 128.0, 126.4, 118.5, 117.7, 91.3, 87.5, 79.3, 79.1, 49.3, 30.8, 28.5, 28.1, 25.5, 21.9, 13.8.

HRMS (MALDI, DCTB): Calculated for $C_{43}H_{49}N_8O_6$ $[M+H]^+$: 773.3770. Found: 773.3767.



CC7. TFA (1.5 mL) was added to a suspension of **CC6** (20 mg, 0.026 mmol) in CH_2Cl_2 (5 mL). The mixture was stirred at room temperature for 29 h. Then, $NaHCO_3$ (sat.) (5 mL) was added and the crude was stirred for 30 min. CH_2Cl_2 was evaporated *in vacuo* and water (5 mL) was added. The suspension was sonicated, filtered and washed with acetonitrile to obtain **CC7** as a yellow solid (16.4 mg, 94%).

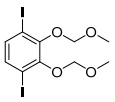
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 8.56 (s, 2H, H^{6C}), 8.15 (bs, 2H, $C^{4C}NH_2$), 7.98 (s, 2H, H^d), 7.35 (d, J = 8.6 Hz, 2H, H^e), 7.12(d, J = 8.3 Hz, 2H, H^f), 3.80 (t, J = 7.4 Hz, 4H, $N^{1C}CH_2C_5H_{11}$), 1.63 (m, 4H, $N^{1C}CH_2CH_2C_4H_9$), 1.27 (m, 12H, $N^{1C}(CH_2)_2(CH_2)_3CH_3$), 0.86 (t, 6H, $N^{1C}C_5H_{10}CH_3$).



GC14. GC14 was prepared following the *Standard Procedure F* using **G1$_{Alk2}$** (16 mg, 0.08 mmol), **G1$_{GC14}$** (43 mg, 0.08 mmol), CuI (0.17 mg, 0.001 mmol) and $Pd(PPh_3)_2Cl_2$ (1.9 mg, 0.003 mmol) in DMF/NEt_3 (15 mL). The reaction was stirred under argon at 60 °C and completed overnight. Then, the reaction mixture was concentrated under reduced pressure and washed with cold acetonitrile, THF and $CHCl_3$/MeOH (100:1) affording **GC14** as a brown solid (29 mg, 57%).

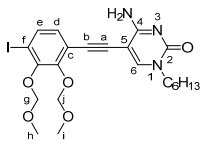
^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.78 (bs, 1H, NH^{1G}), 8.69 (d, J = 5.4 Hz, 1H, H^a), 8.35 (s, 1H, H^{6C}), 8.15–8.04 (m, 2H, H^c, H^e), 7.98 (bs, 1H, $C^{4C}NH$-H), 7.94 (d, J = 8.2 Hz, 1H, H^f), 7.66–7.53 (m, 2H, H^b, H^d), 7.04 (bs, 1H, $C^{4C}NH$-H), 6.72 (bs, 2H, $C^{2G}NH_2$), 4.27 (q, J = 7.3 Hz, 2H, $N^{9G}CH_2CH_3$), 3.74 (t, J = 7.3 Hz, 2H, $N^{1C}CH_2C_5H_{11}$), 1.60 (m, 2H, $N^{1C}CH_2CH_2C_4H_9$), 1.31 (t, J = 7.1 Hz, 3H, $N^{9G}CH_2CH_3$), 1.26 (m, 6H, $N^{1C}(CH_2)_2(CH_2)_3CH_3$), 0.86 (t, J = 6.7 Hz, 3H, $N^{1C}C_5H_{10}CH_3$).

Boronate ester derivatives.



1,4-diiodo-2,3-bis(methoxymethoxy)benzene (B12). A solution of *i*Pr_2NH (2.4 mL, 16.68 mmol) in dry THF (4.6 mL) was added to a solution of **B7** (2.52 g, 6.95 mmol) in dry THF. The mixture was stirred for 30 min at rt. Then, chloro(methoxy)methane (1.3 mL, 16.68 mmol) was added dropwise and the mixture was stirred overnight until completion. The crude mixture was filtered and the solvent was evaporated *in vacuo*. The product was purified by column chromatography on silica gel eluted with $CHCl_3$, affording **B12** as a beige solid (2.07 g, 65%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.27 (s, 2H, Ar), 5.15 (s, 4H, CH_2), 3.67 (s, 6H, OCH_3).

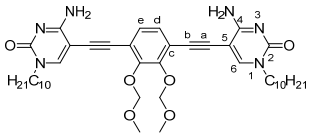


C1$_{GC14}$. C1$_{GC14}$ was prepared following the *Standard Procedure F* using **C2$_{Alk6}$** (0.10 g, 0.46 mmol), diiodoarene **B12** (0.82 g, 1.82 mmol), CuI (1.7 mg, 0.01 mmol) and $Pd(PPh_3)_2Cl_2$ (12.8 mg, 0.02 mmol) in THF/NEt_3 (25 mL). The reaction was completed overnight. Then, the mixture was concentrated under reduced pressure and washed with cold acetonitrile. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (100:1); affording **C1$_{GC12}$** as a brown oil (0.16 g, 64%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.55 (s, 1H, H^6), 7.53 (d, J = 8.4 Hz, 1H, H^e), 6.92 (d, J = 8.3 Hz, 1H, H^d), 6.22 (bs, 2H, NH_2), 5.23 (s, 2H, C^gH_2), 5.16 (s, 2H, C^jH_2), 3.78 (t, J = 7.4 Hz, 2H, $N^1CH_2C_5H_{11}$), 3.67 (s, 3H, OC^hH_3), 3.56 (s, 3H, OC^iH_3), 1.71 (m, 2H, $N^1CH_2CH_2C_4H_9$), 1.30 (m, 6H, $N^1(CH_2)_2(CH_2)_3CH_3$), 0.87 (t, J = 6.3 Hz, 3H, $N^1C_5H_{10}CH_3$).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.4, 154.8, 150.4, 150.2, 147.9, 134.7, 128.9, 118.9, 99.5, 99.5, 93.9, 91.2, 90.2, 85.5, 58.7, 58.2, 50.7, 31.4, 29.2, 26.2, 22.5, 14.0.

HRMS (ESI+): Calculated for $C_{22}H_{29}N_3O_5I$ $[M+H]^+$: 542.1146. Found: 542.1149.

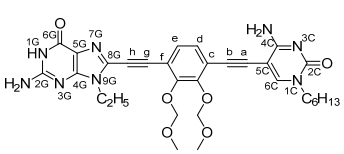


CC8. CC8 was prepared following the *Standard Procedure F* adding a solution of **C2$_{Alk10}$** (0.27 g, 0.98 mmol) in THF/NEt_3 (15 mL) over a mixture of **B12** (0.20 g, 0.44 mmol), CuI (1.0 mg, 0.015 mmol) and $Pd(PPh_3)_2Cl_2$ (10.8 mg, 0.035 mmol) in THF/NEt_3 (15 mL). The reaction was completed overnight. Then, the mixture was concentrated under reduced pressure and washed with cold acetonitrile. The product was purified by column chromatography on silica gel eluted with $CHCl_3$/MeOH (50:1); affording **CC8** as a brown solid (0.27 g, 80%).

^{1}H NMR (300 MHz, $CDCl_3$) δ(ppm) = 7.57 (s, 2H, H^6), 7.15 (s, 2H, H^d), 6.23 (bs, 4H, NH_2), 5.25 (s, 4H, OCH_2O), 3.80 (t, J = 7.2 Hz, 4H, N^{1}CH_2C$_9$H$_{19}$), 3.57 (s, 6H, OCH_3), 1.78–1.67 (m, 4H, N^{1}CH$_2$(CH_2)C$_8$H$_{17}$), 1.34–1.20 (m, 28H, N^1(CH$_2$)$_2$C$_7$$H_{14}CH_3$), 0.87 (t, J = 6.5 Hz, 6H, N^{1}C$_9$H$_{18}$CH_3).

^{13}C NMR (75 MHz, $CDCl_3$) δ(ppm) = 164.5, 154.7, 150.5, 147.9, 127.6, 118.9, 99.4, 91.6, 90.3, 86.3, 58.1, 50.7, 31.8, 29.5, 29.4, 29.22, 29.19, 26.5, 22.6, 14.1.

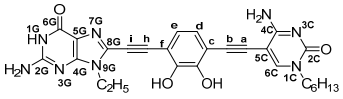
HRMS (MALDI, dithranol): Calculated for $C_{42}H_{61}N_6O_6$ $[M+H]^+$: 745.4647. Found: 767.4472.



GC15. GC15 was prepared following the *Standard Procedure F* using **G1$_{Alk2}$** (63 mg, 0.31 mmol), **C1$_{GC15}$** (0.14 g, 0.26 mmol), CuI (1 mg, 0.005 mmol) and $Pd(PPh_3)_2Cl_2$ (7 mg, 0.01 mmol) in DMF/NEt_3 (15 mL). The reaction was completed overnight. Then, the reaction mixture was concentrated under reduced pressure and washed with THF; $CHCl_3$ and acetonitrile, affording **GC15** as a red solid (148 mg, 93%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.75 (bs, 1H, N^{1G}H), 8.17 (s, 1H, H^{6c}), 7.70 (bs, 1H, C^{4C}NH-H), 7.51 (d, J = 8.1 Hz, 1H, H^e), 7.41 (d, J = 8.3 Hz, 1H, H^d), 6.84 (bs, 1H, C^{4C}NH-H), 6.66 (bs, 2H, C^{2G}NH_2), 5.28 (s, 2H, OCH_2O), 5.26 (s, 2H, OCH_2O), 4.12 (q, J = 7.4 Hz, 2H, N^{9G}CH_2CH$_3$), 3.71 (t, J = 7.4 Hz, 2H, N^{1C}CH_2C$_5$H$_{11}$), 3.55 (s, 3H, OCH_3), 3.53 (s, 3H, OCH_3), 1.59 (m, 2H, N^{1C}CH$_2$CH_2C$_4$H$_9$), 1.34 (t, J = 7.3 Hz, 3H, N^{9G}CH$_2$CH_3), 1.26 (m, 6H, N^{1C}(CH$_2$)$_2$(CH_2)$_3$CH$_3$), 0.86 (t, J = 6.3 Hz, 3H, N^{1C}C$_5$H$_{10}$CH_3).

HRMS (ESI+): Calculated for $C_{31}H_{37}N_8O_6$ $[M+H]^+$: 617.2830. Found: 617.2852.



GC16. HCl. (35%) (0.3 mL) was added to a suspension of **GC16** (43 mg, 0.07 mmol) in a 4:1 $CHCl_3$:MeOH mixture (5 mL) and the mixture was stirred at rt for 18 h. Once completed, $NaHCO_3$ (sat.) (5 mL) was added and the crude mixture was filtered and washed with water; a 10:1 $CHCl_3$/MeOH mixture and acetonitrile, affording **GC16** as a brown solid (25 mg, 69%).

^{1}H NMR (300 MHz, DMSO-d_6) δ(ppm) = 10.72 (bs, 1H, H^{1G}), 9.64 (bs, 2H, OH), 8.18 (s, 1H, H^{6C}), 7.73 (bs, 1H, C^{4C}NH-H), 7.07 (bs, 1H, C^{4C}NH-H), 6.96 (s, 2H, H^e, H^d), 6.62 (bs, 2H, C^{2G}NH_2), 4.12 (m, 2H, N^{9G}CH_2CH$_3$), 3.70 (m, 2H, N^{1C}CH_2C$_5$H$_{11}$), 1.59 (m, 2H, N^{1C}CH$_2$CH_2C$_4$H$_9$), 1.34 (m, 3H, N^{9G}CH$_2$CH_3), 1.26 (m, 6H, N^{1C}(CH$_2$)$_2$(CH_2)$_3$CH$_3$), 0.85 (m, 3H, N^{1C}C$_5$H$_{10}$CH_3).

Summary and Conclusions.

1) Introduction.

In the last decades, supramolecular chemistry has served as a versatile and powerful tool for the construction of highly complex systems and materials through the utilization of different non-covalent interactions that can become much stronger *via* cooperative and multivalent effects. Following a bottom-up strategy and by precisely designing starting single molecules that carry the necessary information (size, shape and functionality), one can engineer self-assembled final structures equipped with intriguing properties.

2) General Objective.

In this Thesis, we have pursued the precise tailoring of new materials by means of the nanostructuration of organic molecular networks on surfaces. Using the tools of molecular self-assembly and the inherent supramolecular properties of DNA nucleobase derivatives, we have sought to achieve a rigorous control at the nanoscale with the aim of transferring the molecular properties to the macroscopic scale. We have focused on the utilization of the extraordinarily selective and directional nucleobase H-bonds for constructing finely-tuned nanoporous 2D networks able to specifically recognize guest molecules as a function of their size, shape and chemical structure.

We have followed a versatile strategy where specific cyclic tetramers have been self-assembled from four ditopic monomeric subunits by Watson-Crick H-bonding interactions. These discrete tetramers were then assembled into ordered porous networks guided by the cooperative action of different secondary interactions (H-bonding, van der Waals forces and interactions with the substrate) (Figure 119).

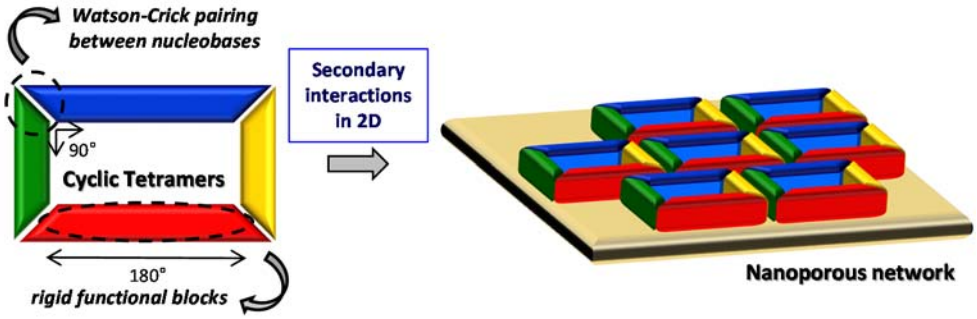
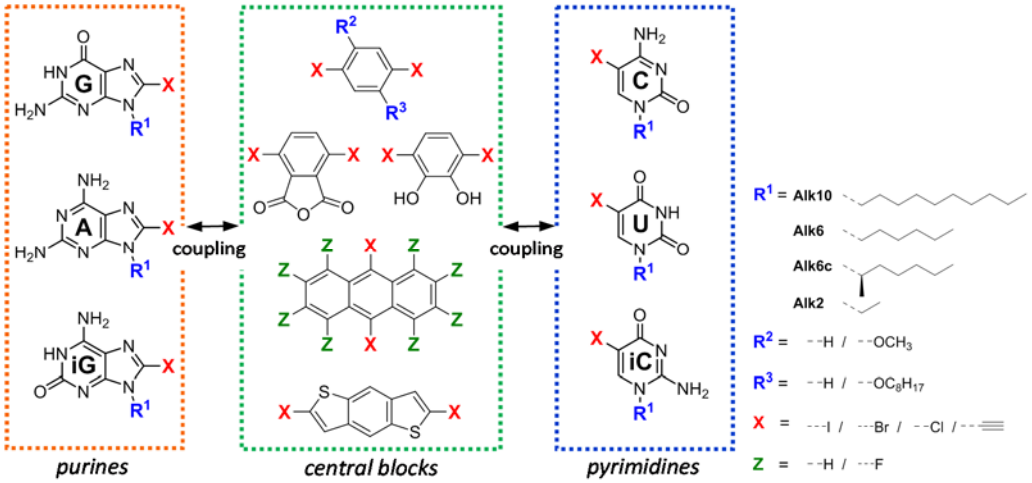


Figure 119. Self-assembly strategy toward nanostructured self-assembled porous networks.

3) Monomer Design and Synthesis.

The target monomers share a common structure based on rigid and planar π-conjugated central blocks that are linearly substituted at both ends with nucleobase directors. Therefore, we have first of all focused on designing and synthesizing the individual molecular components (*Chapter 1*). A convergent synthetic strategy has also

been drawn up in this Chapter in order to reach the final ditopic monomers by conveniently coupling the building blocks. In this first purely synthetic task, a wide family of both the mentioned nucleobase derivatives and dihalogenated central blocks has been synthesized as precursors of final target molecules that will be used for further studies in the next chapters (Scheme 36).



Scheme 36. Molecular components considered and synthesized in Chapter 1.

The chemistry of π-conjugated aromatic molecules that will be used as central blocks has been investigated as well. A useful collection of these molecules fulfilling a number of requirements has been prepared: (1) rigidity; (2) linear disubstitution in opposite directions with halogen atoms (-I, -Br and -Cl); and (3) planarity, so as to afford stability by establishing strong interactions with the substrate. Additionally, some building blocks present interesting electronic properties and some others display specific functionalities reactive toward widely used reversible and non-reversible condensations in covalent organic framework chemistry.

The reactivity of DNA-nucleobase derivatives has been explored; a synthetic route toward appropriately functionalized derivatives has been optimized. The lipophilic nucleobases comprise natural and non-natural derivatives, which are substituted at the 5-(pyrimidines) or 8-position (purines) with either a halogen atom or a terminal triple bond. These include cytosine, *iso*cytosine and uracil as pyrimidine heterocycles, and guanine, *iso*guanine, and 2-aminoadenine as complementary purine bases. The *N*-1/*N*-9 position of these compounds was functionalized with different alkyl groups, well-suited to afford solubility and direct nucleobase assembly. The synthetic sequences leading to the final ethynylated compounds have been optimized for each base attending to convergence, convenience, ease of purification and overall yields. The results indicate that the choice of reaction conditions and the order in which the halogenation, alkylation and Sonogashira

steps are performed is not always trivial and each base required a particular optimized protocol.

4) Self-assembly onto surfaces.

The self-assembly into discrete macrocycles was then studied under different conditions (Figure 120):

(**a**) by STM at the solid-liquid interface after drop-casting deposition of the sample onto HOPG (*Chapter 2*).

(**b**) by STM at the solid-vacuum interface, sublimating the molecules and depositing them on metallic substrates, such as Ag (111) or Au (111), by molecular beam epitaxy (*Chapter 3*).

(**c**) by STM and AFM at the solid-air interface *via* the preparation of Langmuir-Blodgett films at the water-air interface and subsequent transfer to HOPG or Au (111) by vertical dipping (*Chapter 4*).

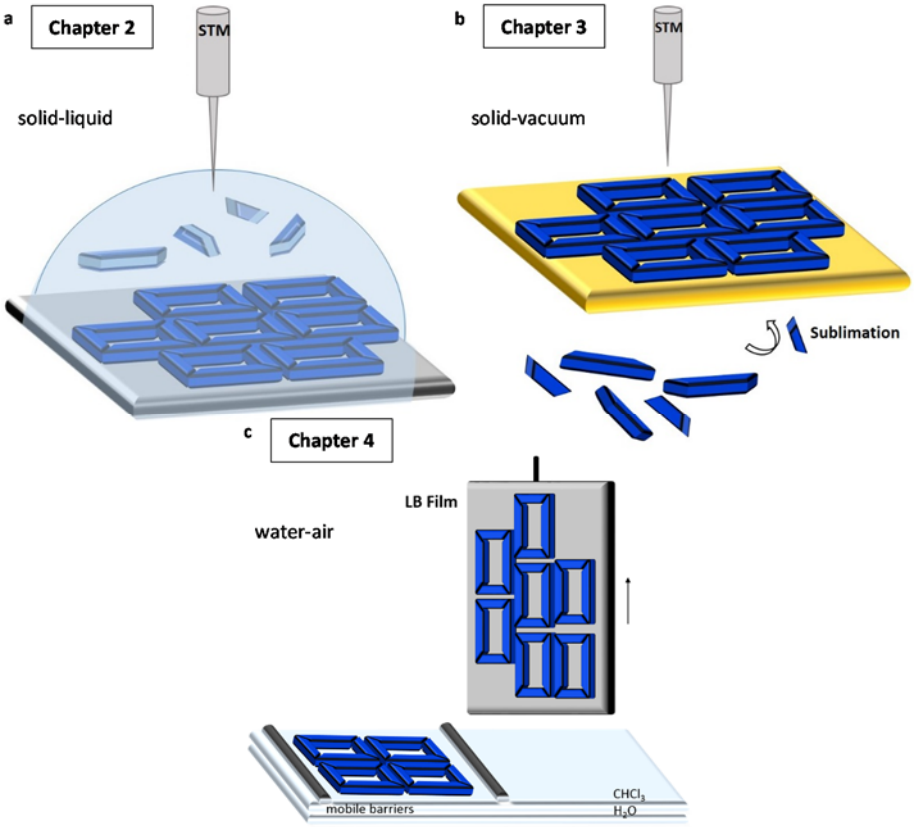


Figure 120. Schematic representation of the self-assembly and imaging strategies approached in this Thesis: (**a**) 2D self-assembly at the solid-liquid interface *via* drop-casting; (**b**) 2D self-assembly at the solid-vacuum interface *via* sublimation; and (**c**) 2D self-assembly at the water-air interface *via* the preparation of Langmuir-Blodgett films.

These three nanostructuring strategies, along with the versatility offered by our monomers, may allow for the study of various subobjectives (Figure 121). However, these have been mainly developed in Chapter 2 at the solid-liquid interface, owing to the fact that this technique is simple, very resourceful, grants a high degree of dynamics and the structural characteristics of the molecules are not as restricted as in other techniques. The stabilization of cyclic tetramer networks on the surface leading to specific nanostructured porous surfaces has been followed as a first objective. A novel approach has been described where instead of building 2D porous networks from weak interactions between di- or tritopic molecules, such as phthalic or trimesic acids, a robust macrocycle was assembled from strong Watson-Crick interactions between DNA bases whose supramolecular structure can be preserved upon surface transfer by rational molecular tailoring. Then, different aspects have been investigated, each presenting a gradual increase in the degree of supramolecular control and in the complexity of the network structure (Figure 121).

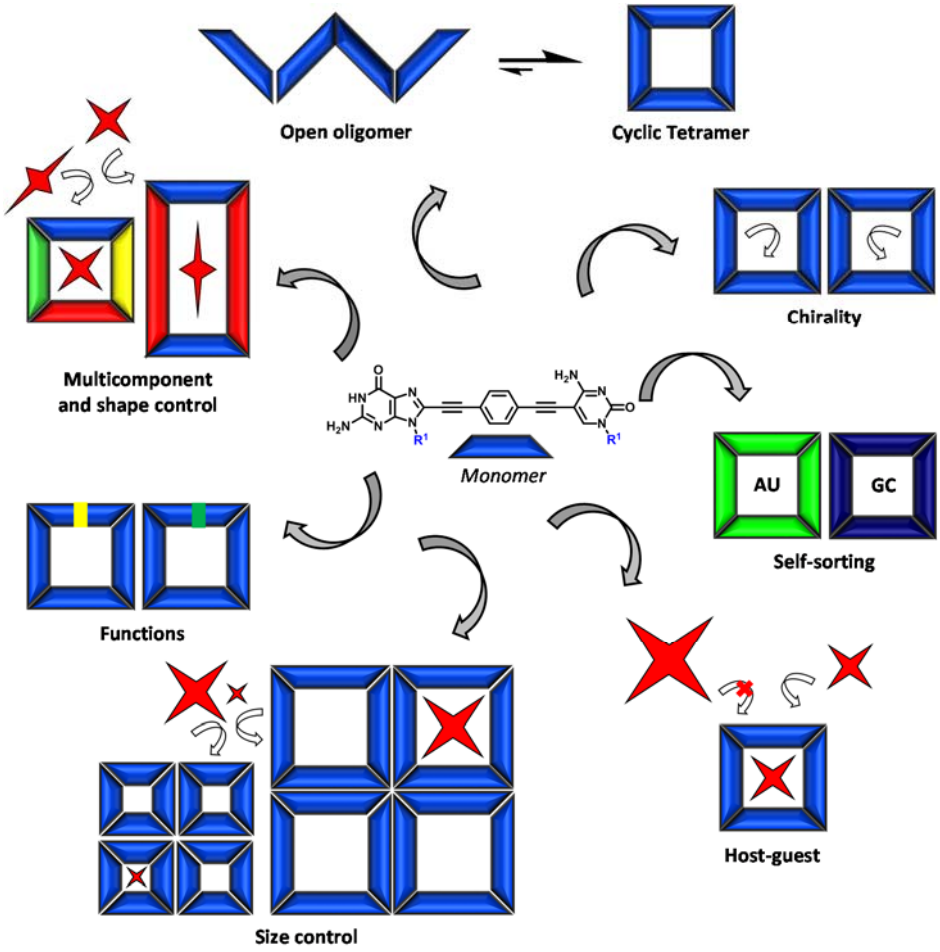


Figure 121. Molecular recognition on nanostructured surfaces: schematic view of different degrees of supramolecular control addressed in this Thesis.

Optimizing cyclic networks. The first task was to stabilize on the surface cyclic tetramers over polymeric species. Even if the molecule is preorganized for ring closure, the intramolecular event that shifts the equilibria toward cyclization do not necessarily apply when the molecules are concentrated on the surface. With this objective in mind, several generations of both GC and AU monomers were designed and synthesized. A rigorous strategy was followed where the experimental results were carefully analyzed and the deduced information was applied to new monomer designs. In this way, stable networks of cyclic tetramers, H-bonded through their Watson-Crick units, were finally obtained. It was seen that the partial occupation of the inner pore of the tetramers with octyloxy chains from the central blocks was important for the stabilization of a dense network of macrocycles. Aside from the Watson-Crick pairs, distinct hierarchical H-bonding secondary interactions were involved in the stabilization of GC and AU tetramer networks. In the GC structures, H-bonding between aminopyridine fragments occurred, so the tetramers were bound through their G edges to form the network (Figure 122a,b). In contrast, in the AU system, secondary H-bonding interactions were established between the external U carbonyl lone pair and A amine proton that were not participating in the Watson-Crick pairing, so the AU tetramers were bound through their corners (Figure 122c,d).

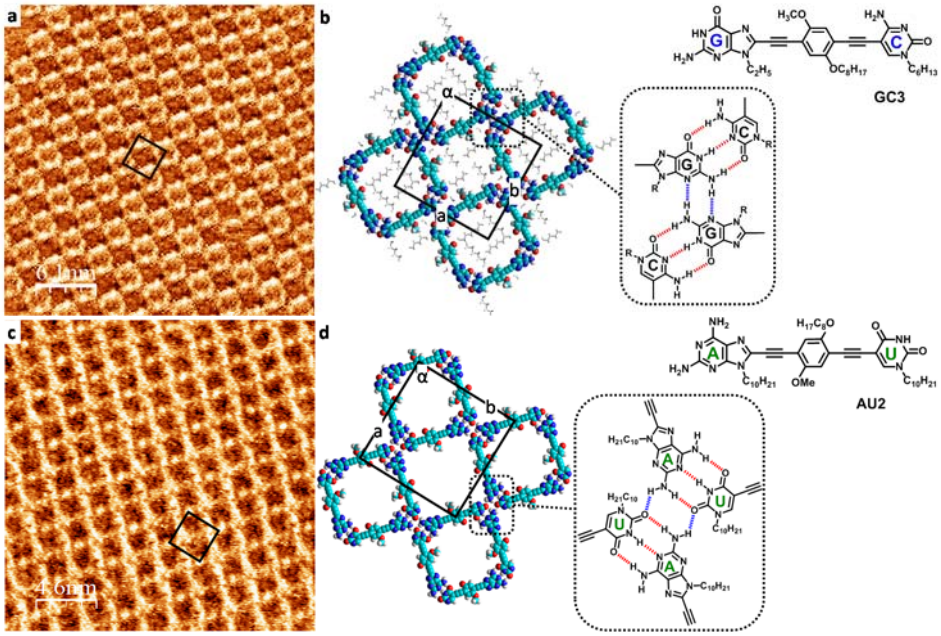


Figure 122. Self-assembled porous networks formed by GC3 and AU2 on HOPG from solutions in TCB:OA (1:1). (**a**) STM image of **GC3** (8.2 x 10^{-6} M; I_{set} = 200 pA, V_{bias} = -300 mV). (**b**) Proposed model for **GC3**, along with chemical structure of the compound and scheme of the stabilizing motif. The unit cell is indicated by black lines (a = 3.6 ±0.1 nm, b = 3.6 ±0.1 nm, α = 89 ±1 °). (**c**) STM image of **AU2** (7.9 x 10^{-5} M; I_{set} = 100 pA, V_{bias} = -350 mV). (**d**) Proposed model for **AU2**, along with chemical structure of the compound and scheme of the stabilizing motif. Unit cell parameters: a = 3.5 ±0.1 nm, b = 3.5 ±0.1 nm, α = 91 ±1°. Dashed red lines indicate Watson-Crick base-pairing and the dashed blue lines correspond to secondary H-bonding between nucleobases.

Host-guest. Inner and outer cyclic tetramer cavities were identified in order to fill with alkyl chains the unspecific space left between macrocycles. The possibility of our persistent cyclic tetramers to incorporate appropriate guest molecules inside their tailored pores was studied at this point. The ability of the nanoporous densely-packed network to host size-complementary guests was proved and stable bimolecular assemblies with coronene (**cor**) were reproducibly obtained (Figure 123). Dual recognition (first between the G and C nucleobases, and then between the cyclic host assembly and the guest molecule) was demonstrated.

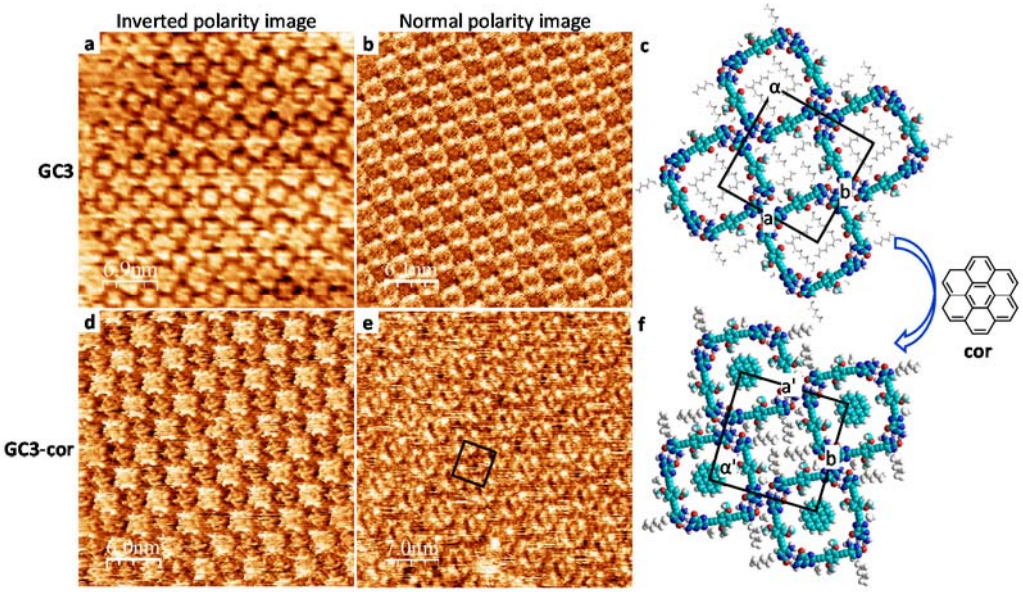


Figure 123. Self-assembly of the GC3 one-component and GC3-cor systems at the HOPG/TCB:OA (1:1) interface. Normal polarity tunneling parameters: I_{set} = 80 pA, V_{bias} = -350 mV. Inverted polarity parameters: I_{set} = 80 pA, V_{bias} = 350 mV. (**a**) Inverted polarity STM image of a single **GC3** domain. (**b**) Normal polarity STM image of **GC3**. (**c**) Model proposal for the one-component system. The unit cell is indicated by black lines (a = 3.6 ±0.1 nm, b = 3.6 ±0.1 nm, α = 89 ±1°). (**d**) Inverted polarity STM image of **GC3** and **cor** (1:200; 6.0 x 10^{-6} M). (**e**) Small scale normal polarity STM image of the **GC3**-**cor** system (**f**) Model proposal for the host-guest system (a' = 3.6 ±0.1 nm, b' = 3.5 ±0.2 nm, α' = 89 ±1°).

Surface chirality. An effort has been made in this Thesis to see how we can control chirality on the surface. Chiral domains have been assembled in *Chapter 2* by the establishment of multiple and distinct H-bonding interactions, of a second-order hierarchy, between tetrameric macrocycles (Figure 124). A chiral monomer **GC5** was prepared to explore the possibility of reaching single pore chirality and thus a chiral network, representing a good starting point toward surfaces potentially able to selectively host particular prochiral guest molecules.

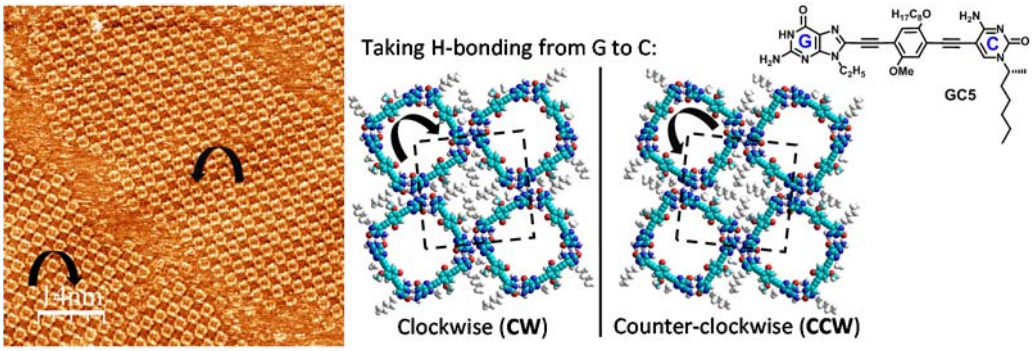


Figure 124. STM image of **GC3** on HOPG from a solution 8.2 x 10^{-6} M in TCB:OA (1:1) (I_{set} = 200 pA, V_{bias} = -300 mV) showing two domains with different pore chirality and explanatory model; along with the chemical structure of the synthesized chiral molecule **GC5**.

Size control. Having obtained complementary host-guest systems with **GC3**, a set of monomers with different sizes was prepared for studying the ability of their 2D H-bonded porous networks to host different guest molecules as a function of their size. Two short GC monomers, a medium-sized one with an unsubstituted central block, and a long GC molecule were synthesized. Well-resolved STM images of the two short molecules corresponding to the formation of cyclic tetramers were also obtained at the solid-liquid interface (Figure 125).

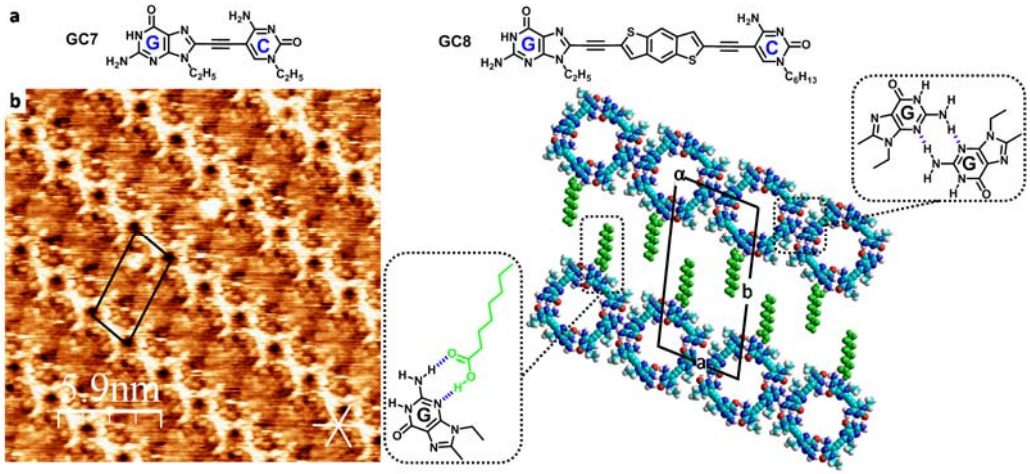


Figure 125. Pore size control and size-selective guest adsorption strategy. (**a**) Chemical structures of the short **GC7** and the long **GC8** monomers. (**b**) STM image of **GC7** self-assembly at the HOPG/TCB:OA (1:1) interface (1.1 x 10^{-6} M; I_{set} = 60 pA, V_{bias} = -300 mV), along with the proposed association model with schematics of the stabilizing H-bonding motifs. The unit cell is indicated by black lines (a = 2.5 ±0.1 nm, b = 4.6 ±0.1 nm, α = 106 ±1°). White lines indicate the normal axes of graphite.

Functions. Monomers with different functional π-electron-rich/-poor blocks have been synthesized as well to explore their effect on the STM images. It was seen that a monomer

presenting an anthracene moiety as central block, displayed a very different electronic density map, compared to other monomers. The molecules presented distinct appearances and were distinguishable on the STM images: while the one carrying a phenyl derivative as central-block was seen as a bright rod, the monomer with the π-electron-rich anthracene appeared as a much brighter dot. Additionally, this molecule displayed two polymorphs: "tetramer" and "hexagon" patterns (Figure 126). The "tetramer" arrangement corresponded to the self-assembly of cyclic tetramers. In the "hexagon" pattern, three **GC9** molecules stablished distorted Watson-Crick bonding and afforded cyclic trimers that were secondarily H-bound through their corners to form the honeycomb structure. Since this monomer is more strongly adsorbed on the HOPG surface by means of its polycyclic aromatic central moiety, the "hexagon" pattern is believed to correspond to a kinetically intermediate situation before reaching the more densely-packed network formed by the "tetramer" pattern. By optimizing distinct experimental parameters, such as time and temperature, which proved to be important factors for the stabilization on the surface of dense 2D arrays of cyclic tetramers, the "tetramer" pattern was favored over the "hexagon" pattern.

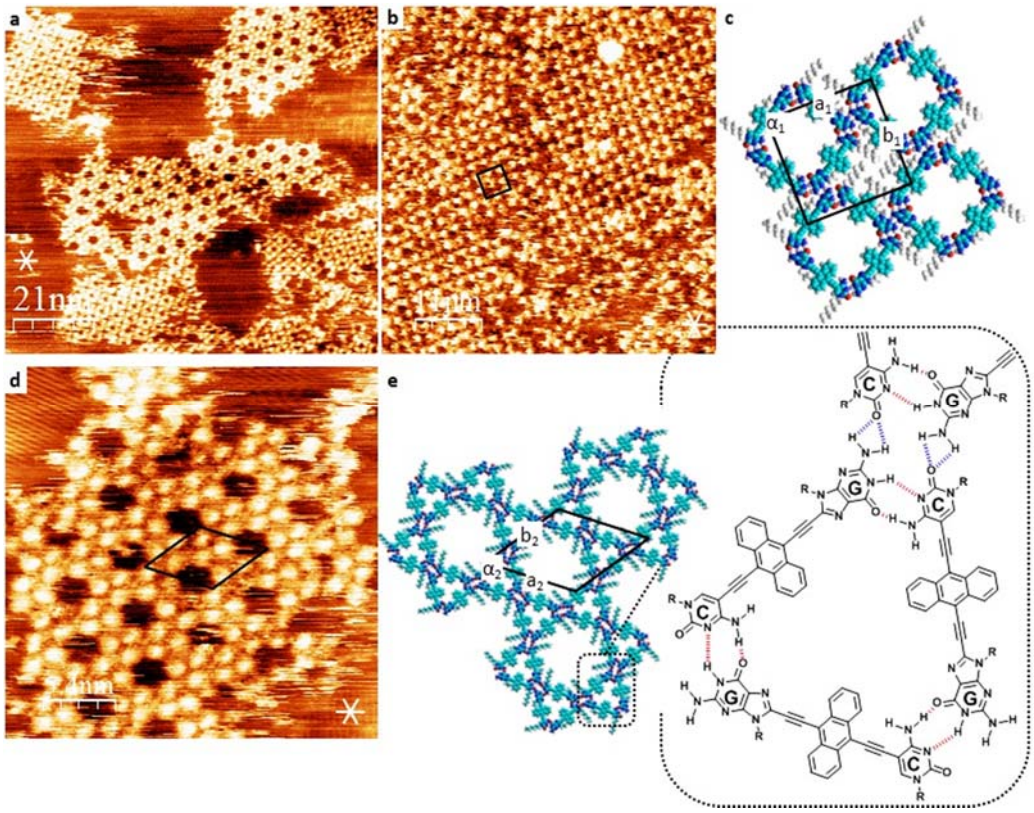


Figure 126. Self-assembly of GC9 on HOPG from solutions in TCB/OA (1:1). White lines indicate the normal axes of graphite. (**a**) STM image of a domain corresponding to the "tetramer" pattern (1.80 x 10^{-5} M; I_{set} = 300 pA, V_{bias} = -100 mV) and (**c**) its association mode, along with the chemical structure of **GC9**.

The unit cell is indicated by black lines (a_1 = 3.4 ±0.1 nm, b_1 = 3.3 ±0.1 nm, α_1 = 90 ±1°). (**c**) STM image of a domain corresponding to the "hexagon" pattern (1.20 x 10^{-5} M; I_{set} = 300 pA, V_{bias} = -150 mV) along with (**d**) proposed association model. Black lines indicate the unit cell (a_2 = 4.6 ±0.1 nm, b_2 = 4.7 ±0.1 nm, α_2 = 60 ±1°).

Self-sorting. The first steps toward self-sorted systems have been given in *Chapter 2*. The specificity and complementarity of the H-bond between DNA-nucleobases was analyzed by co-depositing GC and AU monomers of different length or π-electron density on the substrate and trying to ascertain whether the domains of cyclic tetramers were not mixed and remained formed by only one component (Figure 127). Preliminary studies of the mixture of molecules **AU2** and **GC7** showed that the different nucleobase pairs self-assembled independently into different domains (Figure 127c). Although the experimental conditions have not been optimized yet, the preliminary results were encouraging. Taking into account the differences in the adsorption energy of the single components and their tendency to minimize the occupied areas, control of the deposition of two different molecules on a surface can be considered a difficult task. This approach, preventing the competitive adsorption between the modules, can be considered a valuable route to the future engineering of multicomponent nanostructures on a surface.

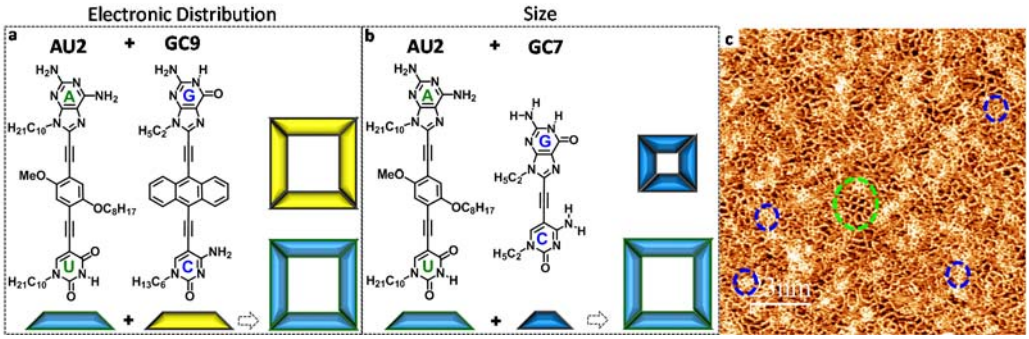


Figure 127. Proposed self-sorting experiments based on (**a**) electronic density and (**b**) size differentiation of our molecular system for the elucidation of self-sorting phenomena on the surface. **GC9** different appearance on the STM images is schematically represented by the yellow color. (**c**) STM image of a 1:3 ratio (**AU2**: 4.0 x 10^{-5} M and **GC7**: 1.2 x 10^{-4} M) premixed solution in TCB/OA (1:1) on HOPG (I_{set} = 80 pA, V_{bias} = -300 mV) corresponding to self-sorting experiment (**b**). Blue dashed circles indicate small GC tetramer domains and green AU bigger macrocycle domain.

Multicomponent and shape control. Multicomponent systems have also been approached in *Chapter 2*. Symmetric monomers of different lengths have been prepared for the construction of two-component cyclic species. Upon association of molecules **GG1** and **CC1**, a rectangular-shaped macrocycle should in principle be formed. Besides, we attempt to investigate shape-discriminating host-guest systems with this network. For instance, we expect an elongated guest molecule, such as pentacene, to be adsorbed within the inner rectangular-shaped pore of the cyclic tetramer. On the other hand, a disk-shaped molecule like **cor** could not be included in this host-guest system (Figure 128).

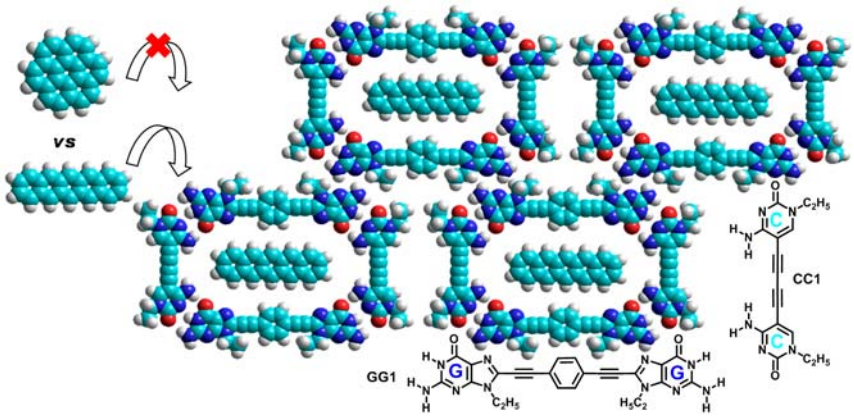


Figure 128. Proposed shape-discrimination experiment between ABAB tetramers (composed of **GG1** and **CC1**), and pentacene as shape complementary guest.

Unfortunately, not all the experiments by STM at the solid-liquid interface have been carried out yet. However, a wide set of molecules has been carefully chosen, prepared, and will be further investigated to complete these studies and pave the way toward molecular recognition at different levels.

The self-assembly process of analogous dinucleobase molecules has been investigated concurrently in *Chapter 3* under vacuum and on different metallic substrates (Au (111); Ag (111)). The experimental conditions were optimized to effectively sublimate and image molecule **GC7** by STM using this technique. A totally-covered gold surface of large domains of a well-resolved 2D porous network was obtained (Figure 129) and a thorough analysis, combining experimental manipulations and theoretical models and calculations in the gas phase, was carried out in order to understand this specific self-assembly process.

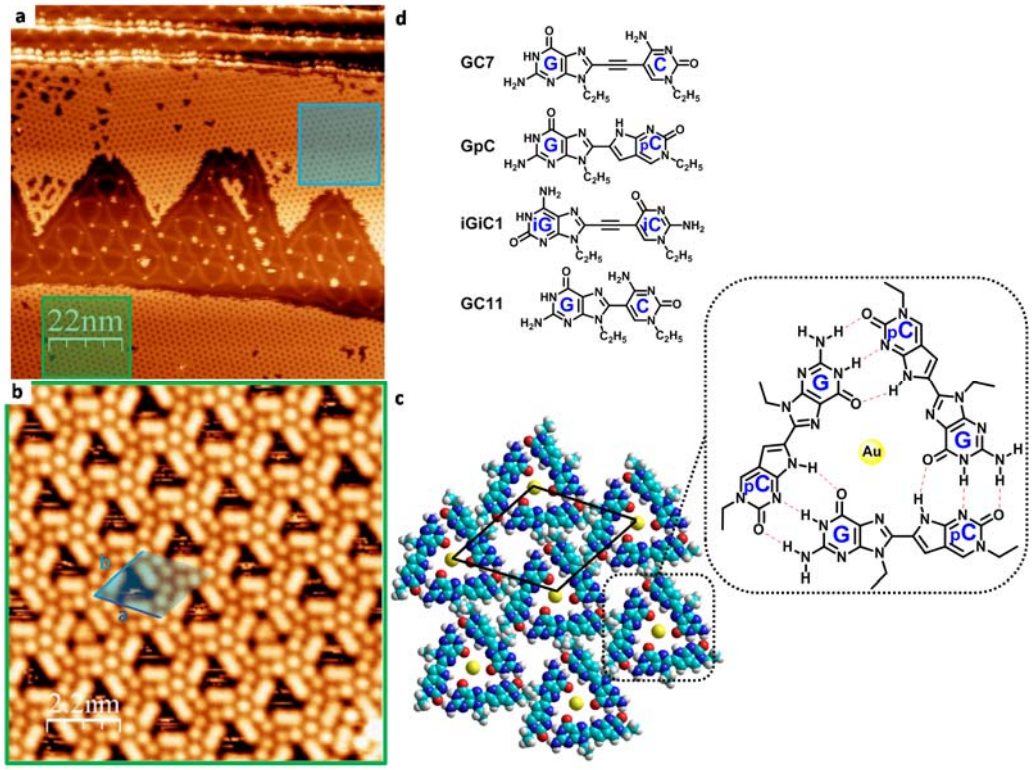


Figure 129. High coverage STM images of GC7 on Au (111). (**a**) Small scale image (tunneling parameters: V_t = -1.5 V; I_t = 5 pA) showing the coexistence of two domains with opposed surface chirality. (**b**) Large scale STM image (V_t = 0.3 V; I_t = 10 pA). The unit cell is marked in blue (a = 2.1 ±0.1 nm; b = 2.0 ±0.1 nm; α = 60 ±0.1°). (**c**) Tentative proposed molecular model considering an H-bonded cyclic trimer based on **GpC** and the stabilization of a Au adatom inside the inner cavity. (**d**) Chemical structure of the considered molecules in Chapter 3.

The STM images showed triangular C_3-symmetric features that could be compatible with an isomerization side-reaction of the molecule **GC7** to form **GpC** (Figure 129d), as experimentally observed after examining the residue left after sublimation. The image interpretation was hampered by the circumstance that our attempts to synthesize **GpC** and/or chemically analyze the sublimation product were unproductive. The best model we could suggest is shown in Figure 129c and considers the adsorption of Au adatoms within the inner pore of a H-bonded cyclic trimer formed by **GpC**. Further experiments will be carried out to achieve the unequivocal model elucidation. Aside from this self-assembly process, different domains displaying different chirality on the surface, corresponding to a different adsorption of the molecular components, were identified (Figure 129a). In addition to the extensive work developed on **GC7** in *Chapter 3*, two novel monomers (**iGiC1** and **GC11**) were carefully designed and their syntheses were approached for the construction of robust self-assembled networks of cyclic tetramers on metallic substrates (Figure 129d). These novel candidates will be studied by UHV-STM in the future.

In *Chapter 4* we concentrated our efforts in reproducing at the water-air interface and by vertical deposition the organization previously obtained with our target monomer **GC3** in Chapter 2. The Langmuir-Blodgett (LB) technique offers the possibility to access, in a controlled manner, materials with a discrete number of layers (bilayer, trilayer and so on). The experimental conditions for the preparation of Langmuir films based on **GC3** at the water-air interface were optimized, as well as the parameters for their transfer onto solid substrates. The addition of ethanol (20%) to solubilize the sample was proven essential to obtain appropriate films at the water-air interface. Otherwise, large aggregates hampered the visualization of good-quality monolayers. The resulting LB films were successfully characterized by AFM and STM techniques and porous monolayers with a structure in agreement with the previous obtained results by STM at the solid-liquid interface were achieved (Figure 130).

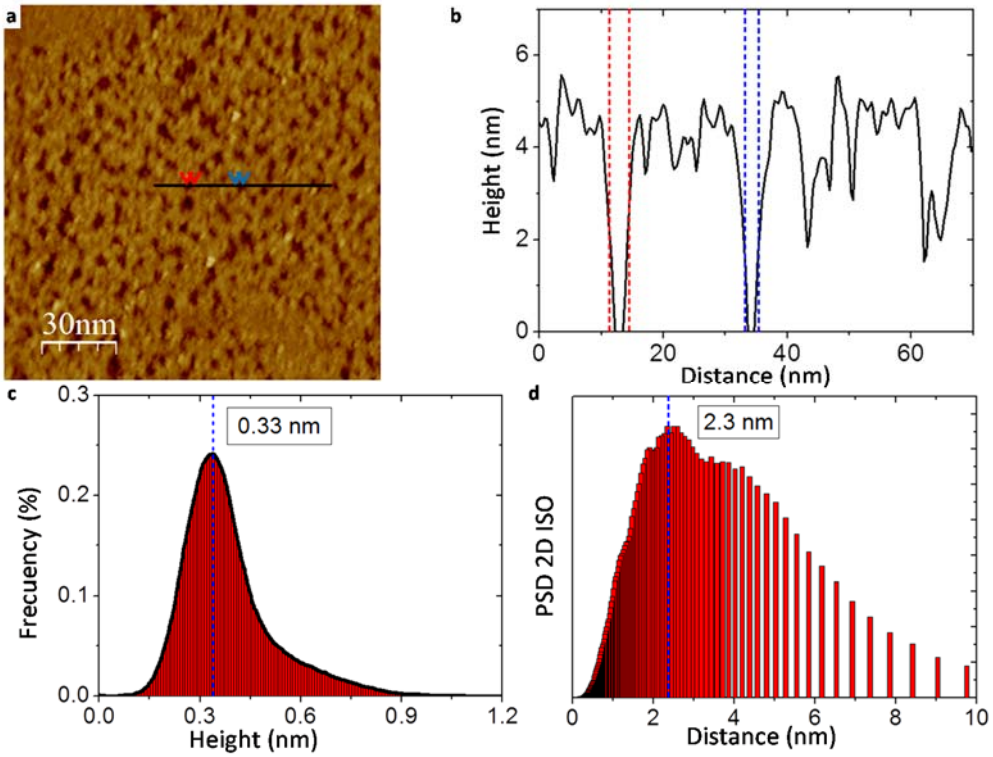


Figure 130. Characterization of a LB film transferred onto Au (111) at 5 mN/m from a monolayer prepared by spreading 750 µL of a **GC3** 4 µM solution in EtOH:$CHCl_3$ (1:4) into water. (**a**) 150 x 150 nm^2 STM image. (**b**) Corresponding typical cross-section analysis. (**c**) Histogram showing the height value distribution confirming the presence of monolayers. (**d**) Isometric 2D power spectral density analysis performed on (**a**), supporting the presence of a porous monolayer with an outer pore of 2.3 nm.

The surface pressure at which the LB films were transferred to solid substrates was proven to be an important factor. High pressures were found to favor the formation of non-

organized multilayered materials where the porous organization could not be resolved. The STM images of the adsorbed layer at low pressures showed an ultrathin porous-shaped organic monolayer whose average height was 0.33 nm, similar to the layer-layer separation on graphite, which is characteristic of monolayers of aromatic moieties placed parallel to the Au (111) surface. The power spectral density analysis exhibited a maximum at 2.3 nm and also a significant major shoulder at 1.9 nm. These results were consistent with the presence of a porous in-plane monolayer with a measured inner pore diameter of 1.9 nm and outer pore diameter of 2.3 nm, which matches the expected size of our cyclic tetramers.

5) Synthesis of fused-monomers for the preparation of H-bonded Organic Frameworks.

A progressive evolution from bidimensional porous networks toward 3D materials has been drawn up in this Thesis. The self-assembly of our molecular building blocks into monolayers has been investigated in Chapters 2 and 3. In Chapter 4, the first steps toward multilayered porous networks have been given. Finally, in *Chapter 5*, covalently fused-monomers have been targeted in order to promote the formation of H-bonded organic frameworks (HOFs) by vertical growth of highly ordered monolayers into crystalline 3D materials (Figure 140). This approach serves as well to guide 2D self-assembly solely relying on strong and directional Watson-Crick interactions (Figure 140, top), and may provide exquisite control over the nature and dimensions of the material cavities.

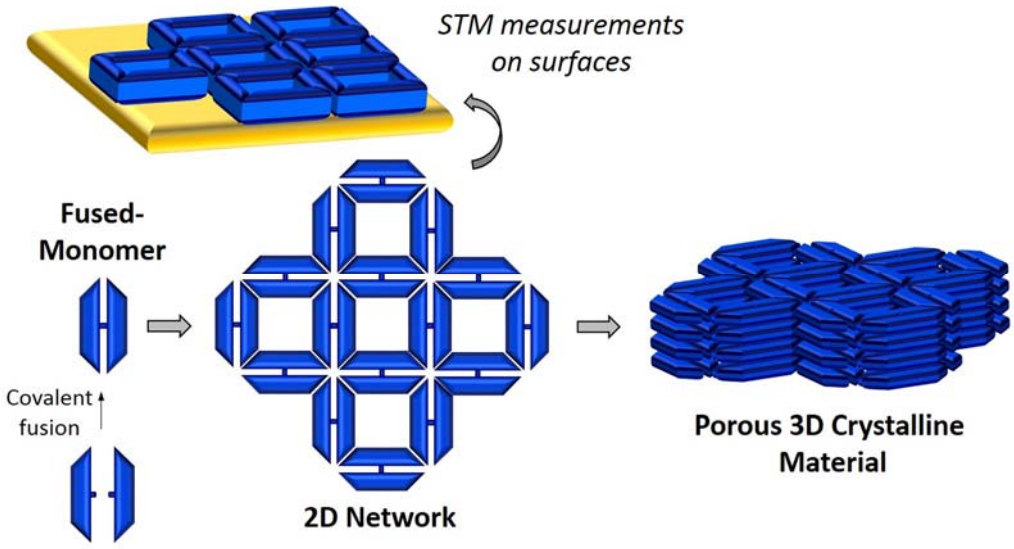


Figure 140. Schematic representation of the objectives of Chapter 5.

Much effort was dedicated to the design of the mentioned HOFs through the construction of several theoretical molecular models. Different symmetric and unsymmetric fused-monomers were considered, with different fusing-motifs (reversible or

irreversible, conformationally flexible or rigid, with different lengths, etc.) but related to our common molecular structure. Two isostructural designs were chosen based on the robust phthalimide linker and the reversible boronic ester union (Figure 141). In the first case, a central block functionalized with an electron-deficient phthalimide derivative was incorporated, so as to afford **GC_GC1** through transimidization reaction with 1,4-diaminobenzene. The second design displayed a diol moiety within the molecule in order to "fuse" two monomers with boronic acid and thus obtain **GC_GC2**. Both fused-monomers have been functionalized with carefully chosen alkyl chains to partially fill the unspecific cavities left between cyclic tetramers. These fused-monomers and other related ones will serve for the preparation of different HOFs, such as chain- or tape-like structures, robust and densely-packed H-bonded bidimensional networks, with square- or rectangular-shaped tailored cavities, and 3D porous crystalline materials formed after stacking of these 2D-sheets.

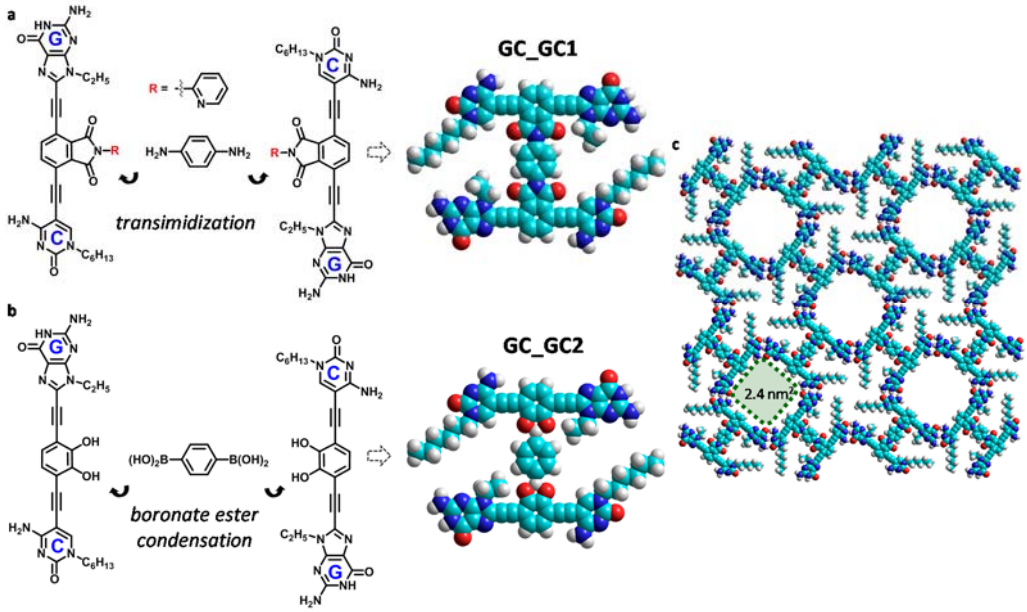


Figure 141. Considered fused-monomers in Chapter 5 for the formation of (**a**) phthalimide- and (**b**) boronate ester-based HOFs. (**c**) Model of the expected 2D layer of porous material where the unspecific secondary pores have been partially blocked with alkyl chains, along with inner cavity dimensions.

In conclusion, a new unconventional and versatile strategy based on molecular self-assembly toward nanostructured surfaces able to specifically recognize molecules as a function of their size, shape and chemical structure has been developed. A large collection of precursors has been synthesized for the construction, *via* Sonogashira coupling reactions, of numerous DNA-based ditopic molecules. These monomers are able to recognize each other and form specific cyclic tetramers that will, in turn, give rise to precisely-tailored nanoporous networks. With these interesting simple molecular building

blocks that can be designed at will, countless experiments can be designed for their investigation under different conditions. We hope that this novel approach will permit to go a step beyond the state-of-the-art in 2D nanostructuring *via* a bottom-up approach and molecular recognition on surfaces.

Resumen y Conclusiones.

1) Introducción.

En los últimos años, la química supramolecular ha servido de herramienta eficaz para construir estructuras complejas, apoyándose en interacciones no covalentes y en su cooperatividad y efectos multivalentes. Siguiendo una estrategia "bottom-up" ('de abajo arriba') y mediante el diseño preciso de las moléculas (tamaño, forma y funcionalidad), se ha dado forma a sofisticados materiales autoensamblados, dotados de propiedades interesantes.

2) Objetivo General.

Esta Tesis Doctoral versa acerca de la nanoestructuración de redes orgánicas sobre superficies para el diseño de nuevos materiales creados a medida. Utilizando bases nitrogenadas presentes en el ADN como andamios supramoleculares, así como las posibilidades que brinda el autoensamblaje molecular, se ha perseguido un control riguroso a nivel nanométrico, para así transferir propiedades específicas a la escala macroscópica. Con este objetivo en mente, hemos aprovechado la extraordinaria selectividad y direccionalidad de los enlaces de hidrógeno para moldear redes bidimensionales nanoporosas, capaces de reconocer moléculas huésped en función de su tamaño, forma y estructura química.

Nos hemos ceñido a una estrategia versátil en la que tetrámeros cíclicos discretos han sido autoensamblados a partir de cuatro unidades monoméricas, mediante interacciones de tipo Watson-Crick. Estos tetrámeros han sido guiados por la acción cooperativa de diversas interacciones secundarias (enlaces de hidrógeno, fuerzas de van der Waals e interacciones con el sustrato sólido) hacia la ordenación en redes porosas (Figura 142).

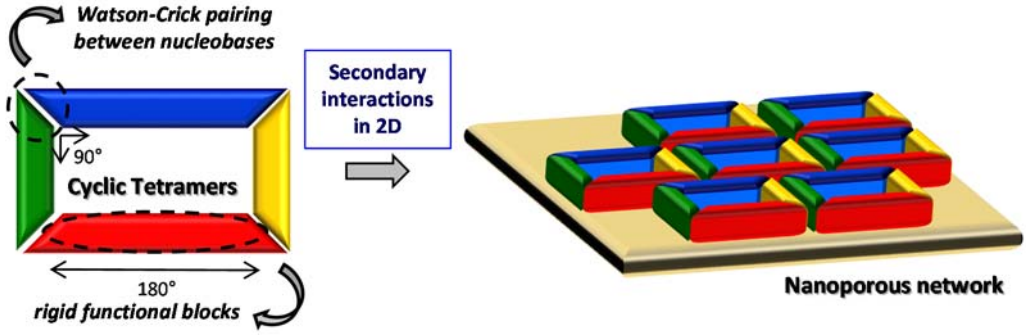
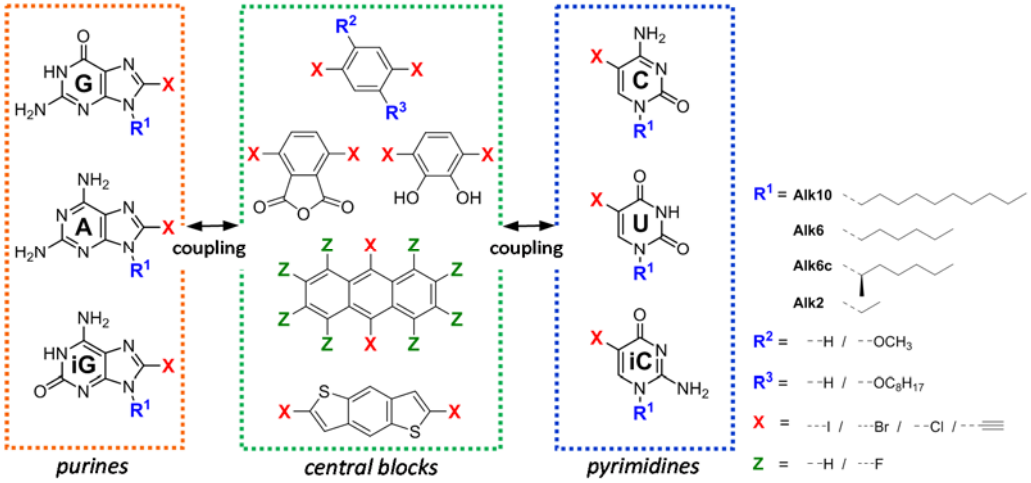


Figura 142. Formación de redes porosas nanoestructuradas mediante el autoensamblaje molecular.

3) Diseño y Síntesis de los Monómeros.

Nuestras moléculas objetivo comparten una estructura común basada en bloques centrales rígidos, planos y π-conjugados, sustituidos en ambos extremos con bases nitrogenadas. En un primer lugar, nos hemos centrado en el diseño y la preparación de cada uno de los componentes (*Capítulo 1*). Se ha trazado una ruta sintética convergente con la que alcanzar estos monómeros finales, mediante el adecuado acoplamiento de las piezas. En esta primera tarea, que es esencialmente sintética, se ha preparado una amplia familia de estos heterociclos y de bloques dihalogenados, que actuarán como precursores de las moléculas objetivo, investigadas en los siguientes capítulos (Esquema 37).



Esquema 37. Componentes moleculares sintetizados en el Capítulo 1.

Se ha investigado la reactividad de moléculas aromáticas π-conjugadas para su uso como bloques centrales y se ha preparado una batería de estos compuestos, teniendo en cuenta requisitos tales como: (1) rigidez; (2) doble sustitución con halógenos (-I, -Br y -Cl), de forma linear y en posiciones opuestas; y (3) planaridad, para que establezcan fuertes interacciones estabilizantes con el sustrato. Además, algunos de los bloques sintetizados presentan propiedades electrónicas interesantes; y otros son reactivos en condensaciones reversibles e irreversibles ampliamente desarrolladas en química de COFs (Covalent Organic Frameworks).

Se ha explorado la reactividad de derivados de bases del ADN, y se han optimizado rutas sintéticas para la funcionalización de estos compuestos con grupos halógeno y etinilo. Estas moléculas engloban derivados naturales y no naturales de bases nitrogenadas que han sido funcionalizadas en la posición 5- (pirimidinas) u 8- (purinas) bien con un halógeno, bien con un triple enlace. Estas incluyen a la citosina, *iso*citosina y uracilo como bases pirimidínicas y a la guanina, *iso*guanina y 2-aminoadenina como heterociclos purina complementarios. Diferentes grupos alquílicos han sido incorporados en la posición *N*-1/*N*-

9 de forma que aporten solubilidad a estos compuestos y dirijan su ensamblado. En el desarrollo de las secuencias sintéticas que permiten llegar a los derivados etinilados, se han valorado elementos tales como convergencia, conveniencia, facilidad de purificación y rendimiento global. Los resultados hallados indican que las condiciones de reacción y el orden en el que se llevan a cabo las etapas de halogenación, alquilación y acoplamiento de Sonogashira no son baladí y cada base nitrogenada requiere de un particular protocolo optimizado.

4) Autoensamblaje sobre superficies.

La formación de macrociclos discretos mediante autoensamblaje se ha estudiado bajo diferentes condiciones (Figura 143).

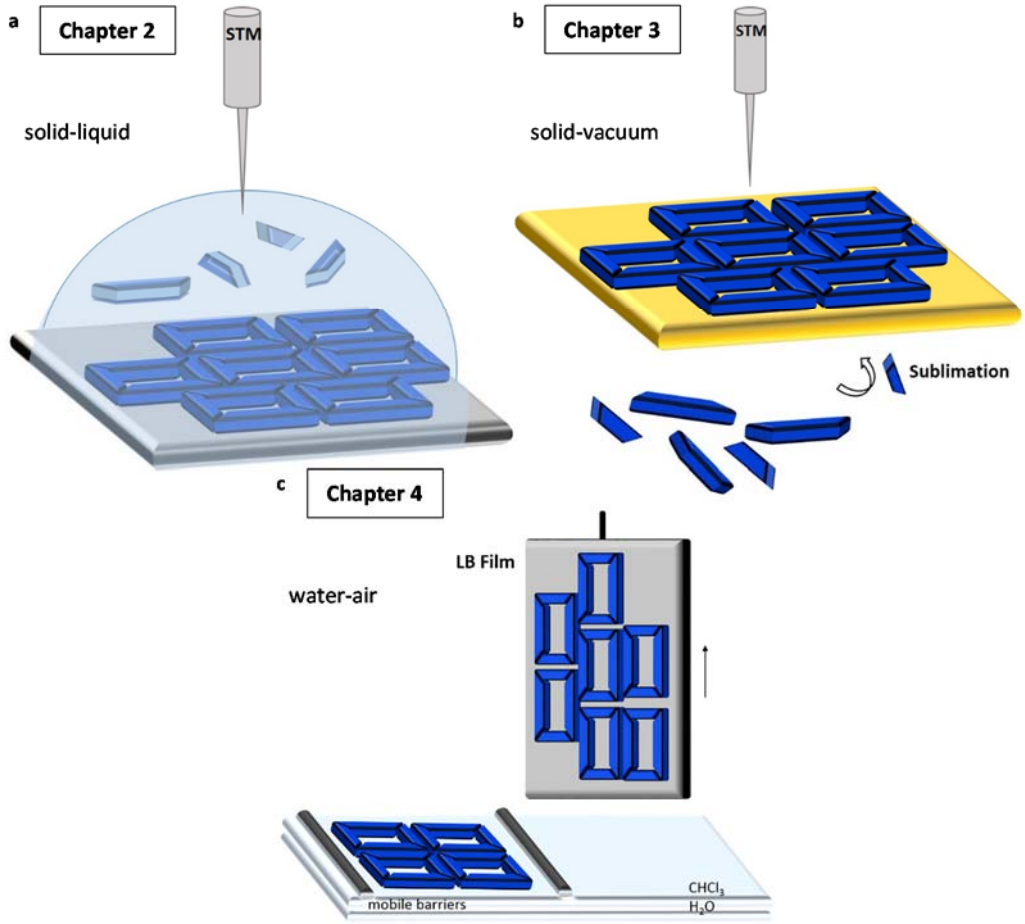


Figura 143. Representación esquemática de las estrategias de autoensamblaje y estudio seguidas en esta Tesis: (**a**) autoensamblaje en la interfase sólido-líquido tras drop-casting; (**b**) autoensamblaje en la interfase sólido-vacío tras sublimación; y (**c**) autoensamblaje en la interfase agua-aire para la preparación de películas de Langmuir-Blodgett.

(**a**) por STM en la interfase sólido-líquido tras depositar la muestra por "drop-casting" (depósito por goteo) sobre HOPG (*Capítulo 2*).
(**b**) por STM en la interfase sólido-vacío por sublimación de la muestra sobre sustratos metálicos, como Ag (111) u Au (111) (*Capítulo 3*).
(**c**) por STM y AFM en la interfase sólido-aire tras la preparación de películas de Langmuir-Blodgett en la interfase agua-aire y posterior deposición por inmersión vertical del sustrato sólido (HOPG u Au (111)) (*Capítulo 4*).

Todas las técnicas de nanoestructuración expuestas previamente, así como la versatilidad que ofrecen nuestros monómeros, permiten el estudio de diversos subobjetivos (Figura 144).

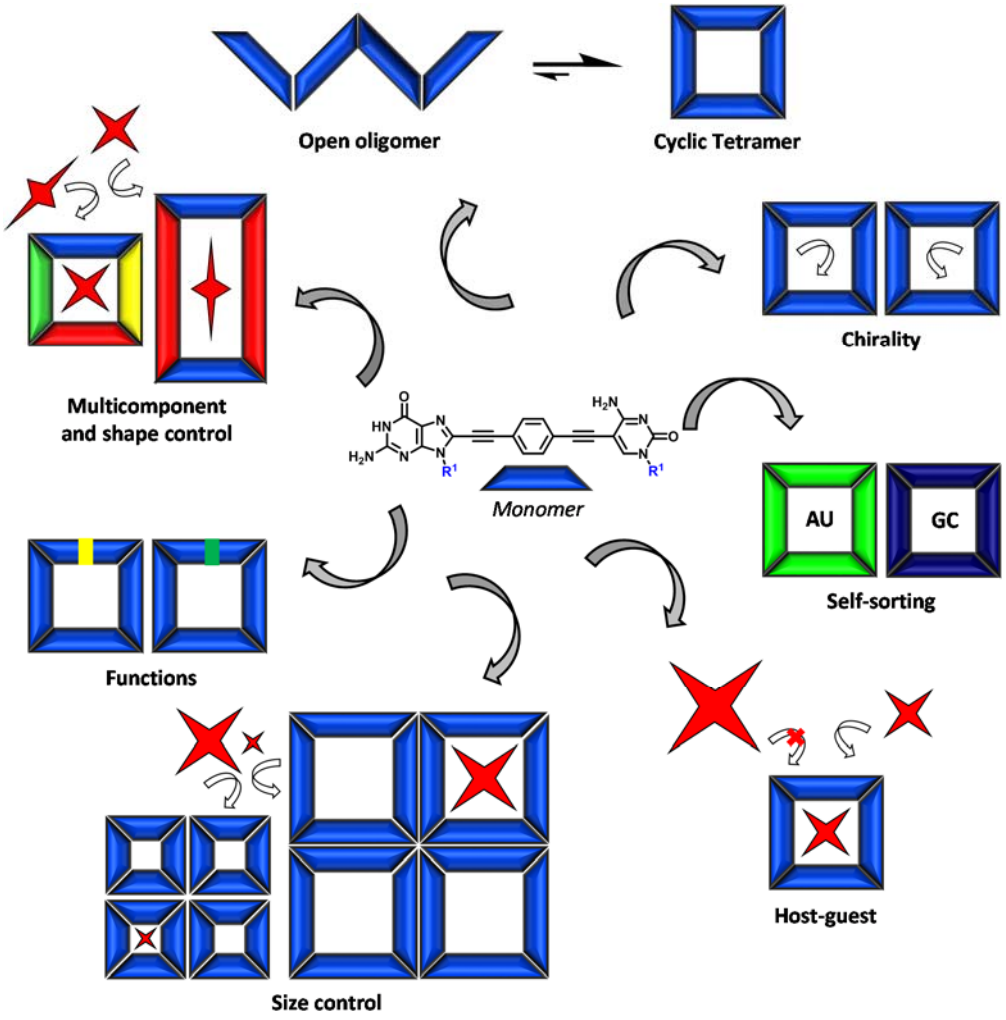


Figura 144. **Reconocimiento molecular sobre superficies nanoestructuradas:** representación esquemática de los distintos grados de control supramolecular considerados en esta Tesis.

Sin embargo, estos han sido desarrollados en su mayoría en el Capítulo 2, por STM en la interfase sólido-líquido, ya que esta metodología combina sencillez, variedad de recursos, características dinámicas y limita la variedad estructural de las moléculas en menor medida, en comparación con otras técnicas. El primer objetivo ha sido la estabilización de redes de tetrámeros cíclicos que conduzcan a la nanoestructuración racional de superficies porosas. Para ello, se ha descrito una aproximación novedosa en la que, en vez de formar redes porosas por interacciones débiles entre moléculas que presentan dos o tres posiciones de anclaje (como el ácido ftálico o el trimésico), se construye un macrociclo robusto a través de interacciones fuertes de tipo Watson-Crick. Gracias a un preciso diseño molecular y al reconocimiento entre bases del ADN, se consigue que esta estructura supramolecular se conserve al transferirla a la superficie. A continuación se han investigado distintos aspectos con un grado de control supramolecular y de complejidad estructural creciente (Figura 144).

Optimización de redes de tetrámeros cíclicos. La primera tarea consistió en favorecer, en la superficie, tetrámeros cíclicos frente a especies de tipo polímero. Aunque una molécula esté preorganizada hacia la formación de un ciclo, este proceso intramolecular que desplaza el equilibrio hacia la ciclación puede no ser relevante cuando las moléculas están concentradas en la superficie. Se ha seguido una estrategia rigurosa en la que los resultados experimentales han sido analizados en detalle para el diseño posterior de nuevos monómeros. Para satisfacer este primer objetivo, se han diseñado y sintetizado varias generaciones de monómeros GC y AU. Finalmente, se pudo obtener redes estables de tetrámeros cíclicos unidos a través de enlaces de hidrógeno entre sus extremos Watson-Crick. Se observó que la ocupación del poro interno del tetrámero con cadenas octiloxi del bloque central era importante para obtener una red densa de macrociclos. Tras el estudio detallado de las interacciones secundarias existentes, se vio que además de las parejas Watson-Crick, otras fuerzas de grado inferior, y que eran diferentes entre tetrámeros GC y AU, intervenían en la estabilización de estas redes. En las estructuras GC, eran enlaces de H entre fragmentos de aminopiridinas los que participaban, de forma que los tetrámeros se unían a través de los lados (Figura 145a,b). Por otro lado, en el sistema AU, los enlaces de H secundarios se establecían entre el par de electrones libre del grupo carbonilo de la U y el protón de la A que no participa en la unión Watson-Crick. De esta manera, estas interacciones acercaban las esquinas de los tetrámeros y daban forma a una red bidimensional con una organización y apariencia diferente a la del sistema GC (Figura 145c,d).

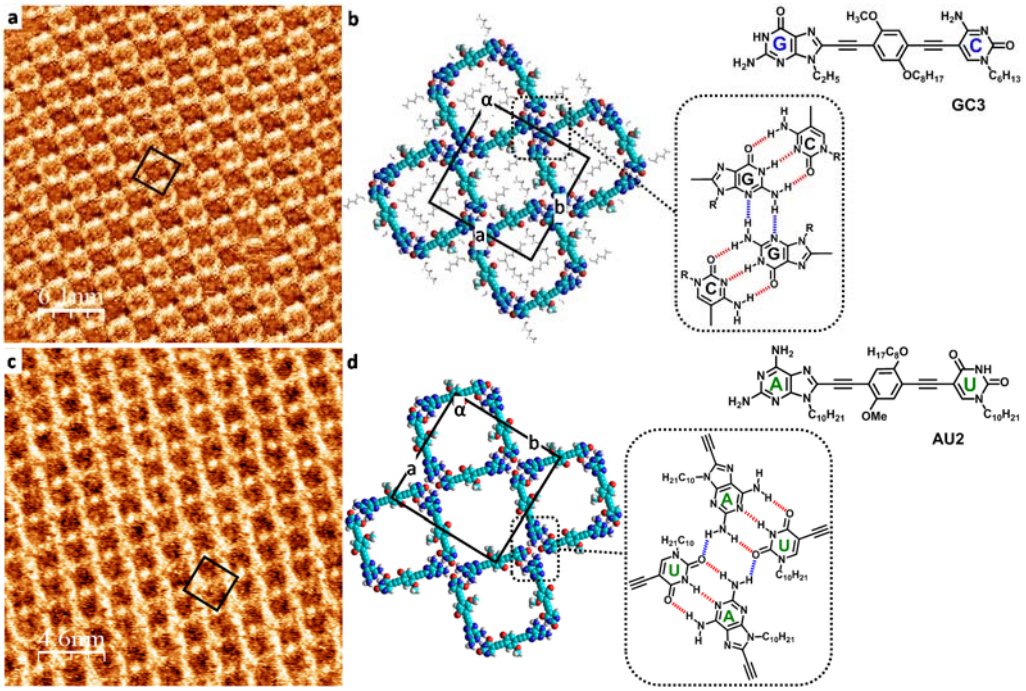


Figura 145. Redes porosas formadas mediante el autoensamblaje de GC3 y AU2 en la interfase HOPG/TCB:OA (1:1). (**a**) Imagen de STM correspondiente a **GC3** (8.2 x 10^{-6} M; I_{set} = 200 pA, V_{bias} = -300 mV). (**b**) Modelo propuesto para **GC3**, junto con su estructura química y un esquema del motivo que estabiliza la red. Las líneas negras indican la celdilla unidad (a = 3.6 ±0.1 nm, b = 3.6 ±0.1 nm, α = 89 ±1 °). (**c**) Imagen de STM correspondiente a **AU2** (7.9 x 10^{-5} M; I_{set} = 100 pA, V_{bias} = -350 mV). (**d**) Modelo propuesto para **AU2**, junto con su estructura y un esquema del motivo que estabiliza la red. Parámetros de la celdilla unidad: a = 3.5 ±0.1 nm, b = 3.5 ±0.1 nm, α = 91 ±1°. Las líneas intermitentes rojas indican el apareamiento Watson-Crick y las azules corresponden a los enlaces de H secundarios entre bases nitrogenadas.

Sistemas host-guest. Se identificaron las cavidades internas y externas de los tetrámeros cíclicos y se ocupó el espacio no específico que quedaba entre macrociclos con cadenas alquílicas. A continuación, se investigó la posibilidad de que estos tetrámeros incorporasen, en el interior de sus poros, moléculas huésped adaptadas. Tras obtener de forma reproducible ensamblajes bimoleculares con coroneno (**cor**), se comprobó que la red nanoporosa previamente formada era capaz de acoger huéspedes de tamaño complementario (Figura 146). De esta forma, se demostró la capacidad de doble reconocimiento molecular de este sistema: primero entre bases G y C; y después entre tetrámero anfitrión y molécula huésped.

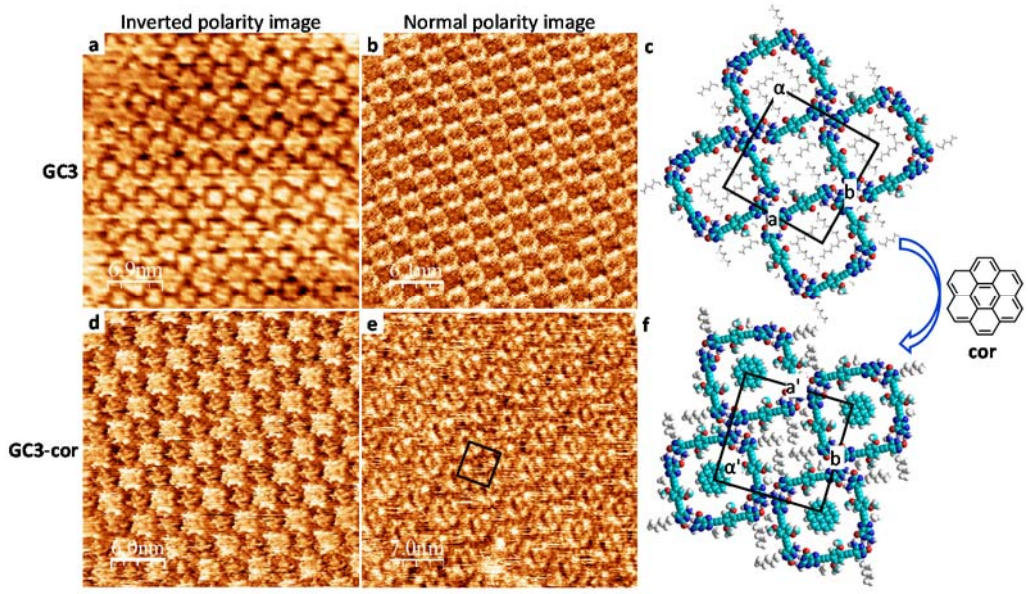


Figure 146. Redes autoensambladas basadas en los sistemas GC3 monocomponente y GC3-cor en la interfase HOPG/TCB:OA (1:1). Parámetros de polaridad normal: I_{set} = 80 pA, V_{bias} = -350 mV. Parámetros de polaridad invertida: I_{set} = 80 pA, V_{bias} = 350 mV. (**a**) Imagen STM de polaridad invertida de un dominio **GC3**. (**b**) Imagen STM de polaridad normal de **GC3**. (**c**) Modelo propuesto para el sistema monocomponente. Las líneas negras indican la celdilla unidad (a = 3.6 ±0.1 nm, b = 3.6 ±0.1 nm, α = 89 ±1°). (**d**) Imagen STM de polaridad invertida del sistema **GC3**-**cor** (1:200; 6.0 x 10^{-6} M). (**e**) Imagen STM de polaridad normal del sistema **GC3**-**cor**. (**f**) Modelo propuesto para el sistema host-guest (a' = 3.6 ±0.1 nm, b' = 3.5 ±0.2 nm, α' = 89 ±1°).

Quiralidad en superficie. En esta Tesis se ha dedicado un esfuerzo en ver cómo se puede controlar la quiralidad en la superficie a través de nuestro sistema. En el *Capítulo 2*, se han ensamblado dominios quirales a través del establecimiento de múltiples y distintas interacciones de tipo enlace de H de un segundo orden entre tetrámeros cíclicos (Figura 147). Se ha sintetizado un monómero quiral (**GC5**) para explorar la posibilidad de alcanzar redes quirales y con ellas, superficies con la capacidad potencial de reconocer selectivamente moléculas huésped proquirales.

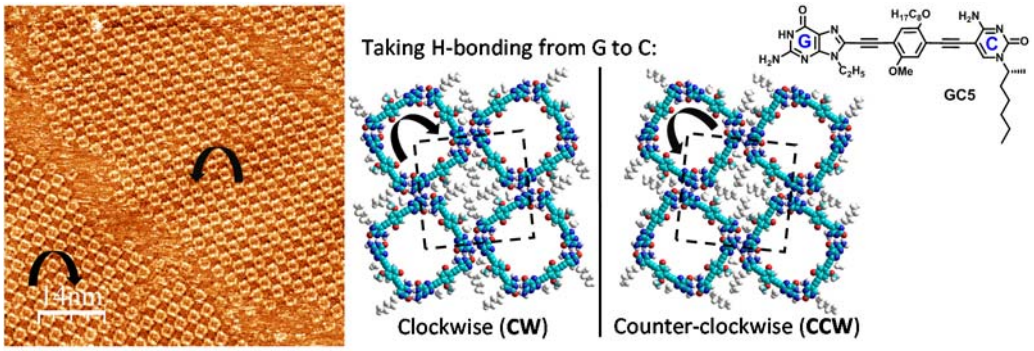


Figura 147. Imagen STM de GC3 en la interfase HOPG/TCB:OA (1:1) (8.2 x 10^{-6} M; I_{set} = 200 pA, V_{bias} = -300 mV) que muestra dos dominios con poros de diferente quiralidad. Modelo explicativo y estructura química de la molécula quiral **GC5**.

Control por tamaño. Habiendo obtenido sistemas host-guest complementarios con **GC3**, se preparó una serie de monómeros de diferente longitud para estudiar la formación de redes porosas capaces de acoger moléculas huésped en función de su tamaño. Con este objetivo, se sintetizaron dos monómeros GC cortos; uno de longitud media, como **GC3**, pero cuyo bloque central no estaba sustituido; y un monómero GC largo. Las moléculas cortas se estudiaron por STM en la interfase sólido-líquido y permitieron la visualización de tetrámeros cíclicos (Figura 148). El resto de productos será estudiado más adelante.

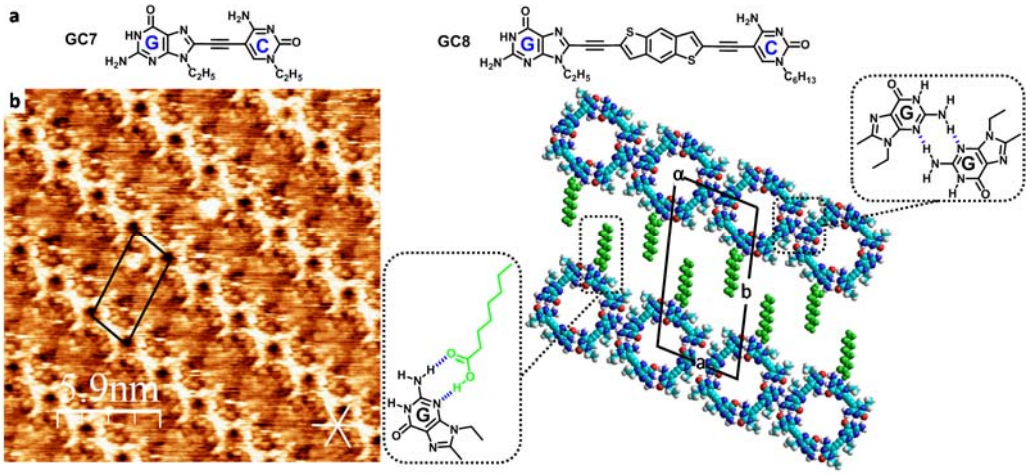


Figura 148. Experimentos para controlar el tamaño del poro y la adsorción selectiva de huéspedes. (**a**) Estructuras de los monómeros **GC7** y **GC8**. (**b**) Imagen STM de **GC7** en la interfase HOPG/TCB:OA (1:1) (1.1 x 10^{-6} M; I_{set} = 60 pA, V_{bias} = -300 mV), junto con el modo de asociación propuesto y esquemas de los motivos que estabilizan la red. Las líneas negras indican la celdilla unidad (a = 2.5 ±0.1 nm, b = 4.6 ±0.1 nm, α = 106 ±1°). Las líneas blancas indican los ejes normales del grafito.

Función. Por otro lado, también se han preparado monómeros dotados de bloques ricos y pobres en electrones, para con ellos explorar su efecto sobre las imágenes de STM. En el

curso de esta investigación, se ha visto que un monómero funcionalizado con un grupo antraceno daba lugar a un mapa de densidad electrónica muy dispar, en comparación a otras moléculas previamente estudiadas. Mientras que los compuestos dotados de un derivado bencénico como grupo central aparecían en las imágenes de STM como una varilla de brillo moderado, los monómeros que llevaban por bloque central un grupo rico en electrones, como el antraceno, se mostraban como un disco muy brillante. Por otro lado, esta molécula presentaba dos patrones diferentes (Figura 149). El primer motivo ha sido denominado "tetrámero" y corresponde efectivamente a los macrociclos deseados. La segunda estructura, llamada "hexágono", corresponde a una situación cinética intermedia, previa a la formación de una red más densa de tetrámeros cíclicos. Las bases nitrogenadas establecen enlaces distorsionados de tipo Watson-Crick y así **GC9** forma trímeros cíclicos y estos se asocian mediante enlaces de H para dar un red muy porosa. Su origen puede ser la mayor afinidad por el sustrato de esta especie policíclica aromática, que se traduce en una menor movilidad de las moléculas sobre la superficie. Al optimizar diversos parámetros experimentales, como la temperatura o el tiempo de espera previo al análisis de la muestra, se pudo favorecer el motivo "tetrámero", termodinámicamente más estable, frente a los dominios de tipo "hexágono".

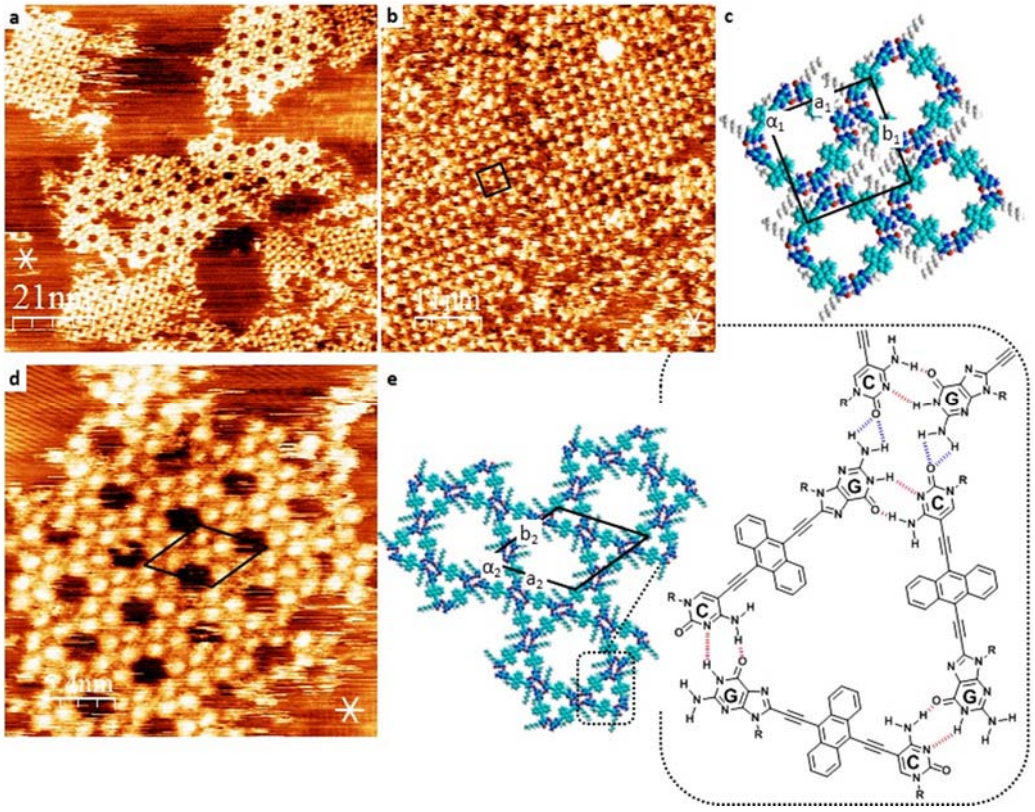


Figura 149. Autoensamblaje de GC9 en la interfase HOPG/TCB:OA (1:1). Las líneas blancas indican los ejes normales del grafito. (**a**) Imagen de STM correspondiente al motivo "tetrámero" (1.80 x 10^{-5} M; I_{set} =

300 pA, V_{bias} = -100 mV) y (**c**) su modo de asociación, junto con la estructura de **GC9**. Las líneas negras indican la celdilla unidad (a_1 = 3.4 ±0.1 nm, b_1 = 3.3 ±0.1 nm, α_1 = 90 ±1°). (**c**) Imagen STM de un dominio de tipo "hexágono" (1.20 x 10^{-5} M; I_{set} = 300 pA, V_{bias} = -150 mV) junto con (**d**) el modelo de asociación propuesto. Dimensiones de la celdilla unidad: a_2 = 4.6 ±0.1 nm, b_2 = 4.7 ±0.1 nm, α_2 = 60 ±1°.

Self-sorting. En el *Capítulo 2* se ha avanzado hacia la creación de sistemas de tipo "self-sorting" basados en la complementariedad del enlace de H entre bases del ADN. Se ha evaluado que, tras depositar una mezcla de moléculas GC y AU de distinta longitud o de diferente densidad π-electrónica, estas formen tetrámeros puros compuestos por un único componente. Para ello, se han llevado a cabo estudios preliminares de mezclas de **AU2** y **GC7**, donde se ha visto que cada una de las moléculas se asocia de forma independiente, y que por lo tanto han resultado prometedores (Figura 150).

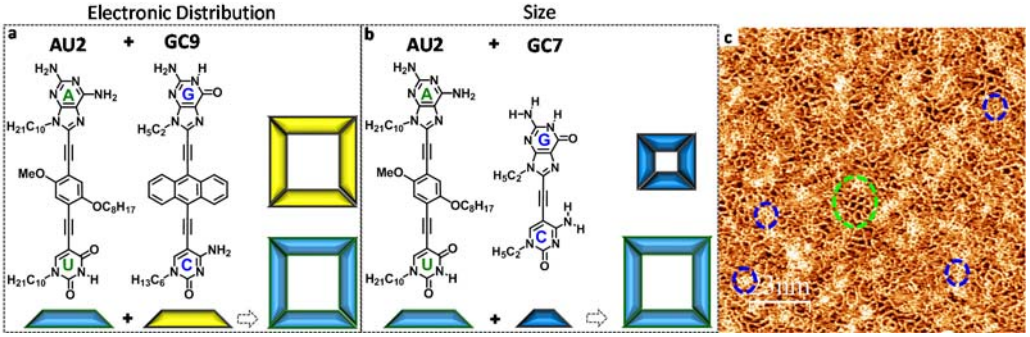


Figura 150. Experimentos de "self-sorting" propuestos basados en (**a**) la diferente densidad electrónica y (**b**) la diferenciación por tamaño. Se ha representado la distinta apariencia de **GC9** en las imágenes de STM con el color amarillo. (**c**) Imagen de STM image de una mezcla 1:3 (**AU2**: 4.0 x 10^{-5} M and **GC7**: 1.2 x 10^{-4} M) en TCB:OA (1:1) sobre HOPG (I_{set} = 80 pA, V_{bias} = -300 mV) correspondiente al experimento (**b**). Los círculos azules indican tetrámeros pequeños GC y el círculo verde un dominio de macrociclos AU de mayor tamaño.

Sistemas multicomponente y control por forma. En el *Capítulo 2* también se ha tratado de estudiar sistemas multicomponente como otro subobjetivo. Para ello, se han sintetizado monómeros simétricos de diferente longitud para la construcción de especies cíclicas formadas por dos componentes. Al asociarse **GG1** con **CC1**, se formaría un macrociclo rectangular con el que buscaremos investigar la adsorción selectiva de moléculas huésped en función de su forma. Por ejemplo, un huésped alargado como el pentaceno podría ser reconocido por una red anfitriona nanoporosa de tetrámeros formados por **GG1** y **CC1**. Sin embargo, una molécula con forma circular como el **cor** no podría ser incluido en esta red de poros rectangulares (Figura 151).

Desafortunadamente, no todos los experimentos de STM en la interfase sólido-líquido han podido llevarse a cabo en el lapso de esta Tesis. Sin embargo, se ha diseñado y preparado una amplia colección de moléculas que serán investigadas próximamente y con las que buscamos demostrar fenómenos de reconocimiento molecular a diferentes niveles.

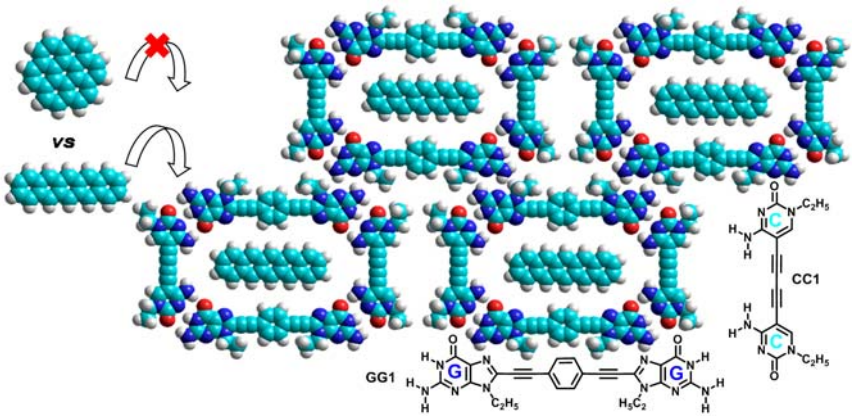


Figura 151. Experimento para investigar la discriminación de moléculas huésped en función de su forma. Se propone una red anfitriona de tetrámeros de tipo ABAB (formados por **GG1** y **CC1**) y al pentaceno como huésped.

De forma paralela, el autoensamblaje de moléculas análogas se ha investigado por STM en ultra alto vacío y sobre sustratos metálicos. Se han optimizado las condiciones experimentales para sublimar una molécula **GC7** y visualizarla en estas condiciones. Se obtuvo con gran resolución una red porosa bidimensional de grandes dominios que recubrían casi por completo la superficie de oro (Figura 152). A lo largo de este Capítulo se ha realizado un estudio exhaustivo para la comprensión de este proceso de autoensamblaje. Para ello se han combinado procedimientos experimentales, modelos teóricos y cálculos en fase gas.

Las imágenes de STM mostraron formas triangulares de simetría C_3 que podrían ser compatibles con una reacción secundaria de isomerización que afectaría a **GC7** y daría lugar a **GpC** (Figura 152d), como se observó experimentalmente al examinar el residuo tras sublimación. La interpretación de las imágenes de STM se vio dificultada al no poder sintetizar **GpC** y/o analizar químicamente el producto de sublimación. El mejor modelo al que se pudo llegar se muestra en la Figura 152c y considera la adsorción de átomos de Au en el interior de tetrámeros cíclicos unidos por 9 enlaces de H y formados por **GpC**. El estudio para el esclarecimiento del modelo molecular sigue llevándose a cabo y más experimentos se realizarán próximamente con este objetivo. Por otro lado, se identificaron dominios de quiralidad opuesta en la superficie, correspondientes a la distinta adsorción de los componentes moleculares (Figura 152a). Además del proyecto desarrollado sobre la molécula **GC7**, en el *Capítulo 3* también se diseñaron dos nuevos monómeros (**iGiC1** y **GC11**) para la creación de redes autoensambladas de tetrámeros cíclicos sobre superficies metálicas (Figura 152e). Estos productos serán estudiados por UHV-STM en adelante.

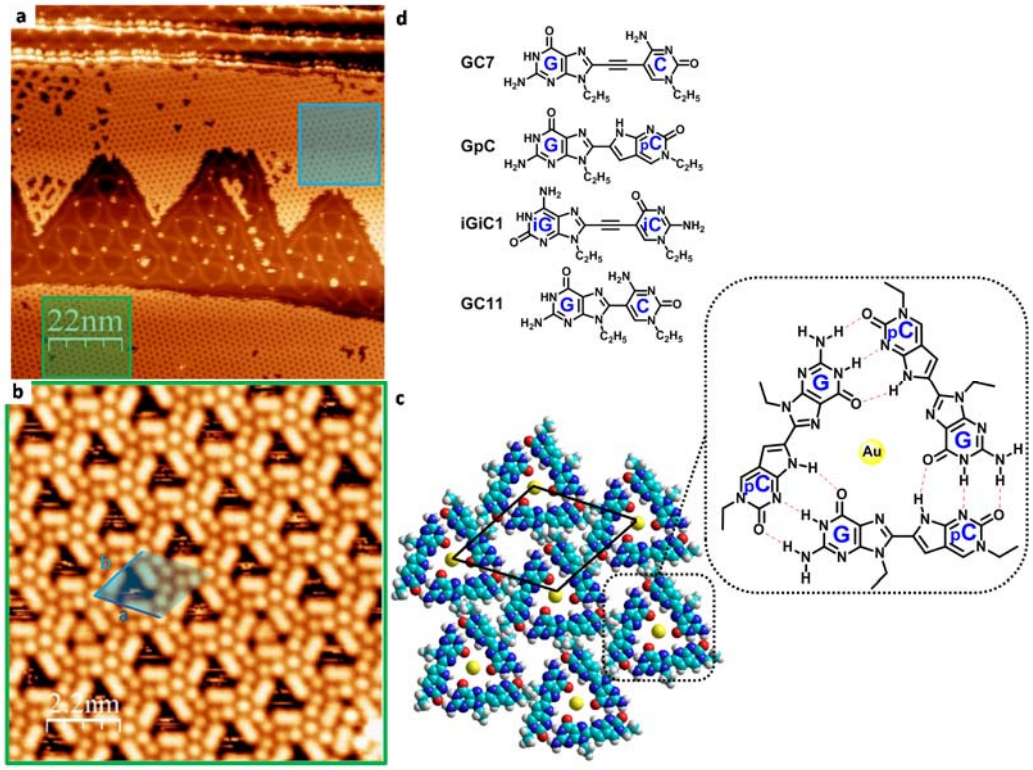


Figura 152. Imágenes de alto recubrimiento de GC7 sobre Au (111). (**a**) Imagen a pequeña escala (V_t = -1.5 V; I_t = 5 pA) que muestra la coexistencia de dos dominios con quiralidad en superficie opuesta. (**b**) Imagen a gran escala (V_t = 0.3 V; I_t = 10 pA). Las líneas azules indican la celdilla unidad (a = 2.1 ±0.1 nm; b = 2.0 ±0.1 nm; α = 60 ±0.1°). (**c**) Modelo propuesto que considera la formación de un trímero cíclico unido por enlaces de H basado en **GpC** y la estabilización de un átomo de Au adsorbido en el interior del poro del trímero. (**d**) Estructuras de los compuestos considerados en el Capítulo 3.

En el *Capítulo 4*, se trató de reproducir los resultados previamente obtenidos con **GC3**, pero esta vez en la interfase agua-aire y por deposición por inmersión del sustrato sólido. La técnica de Langmuir-Blodgett permite acceder de manera controlada a películas con un espesor muy preciso, que depende del número de ciclos de inmersión (o pasos de fuga) que se han llevado a cabo. Se optimizaron las condiciones experimentales para la preparación de películas de Langmuir sobre la superficie del agua, así como los parámetros usados en el proceso de transferencia de la película al sustrato sólido. El uso de etanol (20%) resultó ser determinante para solubilizar la muestra en cloroformo y así obtener películas adecuadas en la interfase agua-aire. De no proceder de esta manera, la formación de agregados voluminosos impedía la visualización de monocapas de buena calidad. Las películas de Langmuir-Blodgett fueron caracterizadas con éxito tanto por AFM como por STM y las imágenes correspondientes a monocapas mostraron una estructura porosa que concordaba con los resultados recabados con anterioridad por STM en la interfase sólido-líquido (Figura 153).

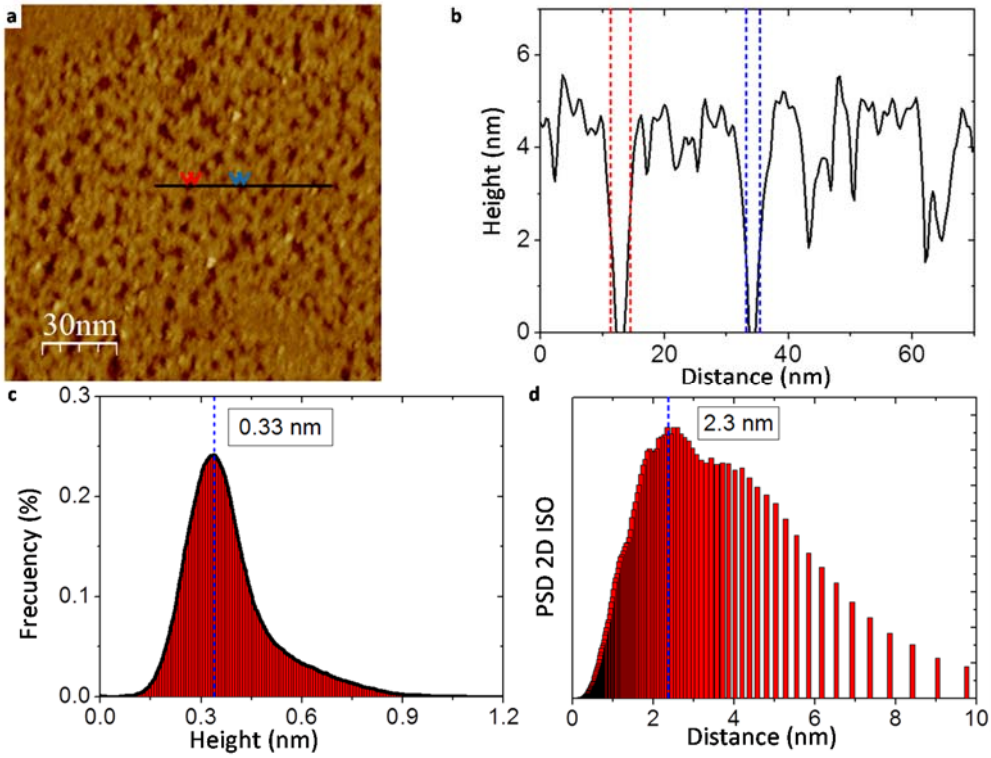


Figura 153. Caracterización de una película LB transferida sobre Au (111) a 5 mN/m a partir de una monocapa preparada tras rociar sobre agua 750 µL de una disolución 4 µM de **GC3** en EtOH:$CHCl_3$ (1:4). (**a**) Imagen de STM de la película LB (150 x 150 nm^2). (**b**) Análisis transversal correspondiente. (**c**) Histograma que muestra la distribución de las diferentes alturas registradas y que confirma la presencia de una monocapa. (**d**) Análisis de la densidad espectral de rugosidad por "Isometric 2D Power Spectral Density" llevado a cabo sobre la imagen (**a**) y que respalda la existencia de una monocapa porosa con una medida de poro externo de 2.3 nm.

La presión superficial a la que se transfiere la película LB ha sido determinante para la obtención de monocapas. Un valor demasiado elevado de este parámetro experimental, se traducía en la formación de materiales multicapa no organizados. Por otro lado, se logró resolver por STM la organización porosa de una monocapa de un grosor de 0.33 nm, que es característica de la separación entre láminas de carbono en el grafito y que concuerda con una monocapa aromática paralela a la superficie de Au (111). El análisis de la densidad espectral de rugosidad puso de manifiesto unos valores máximos de 2.3 nm y 1.9 nm. Estas medidas encajan con la existencia de una monocapa definida por estos diámetros respectivos de poro externo e interno; y que a su vez se ajustan a las dimensiones de nuestro tetrámero cíclico.

5) Síntesis de monómeros fusionados para la preparación de estructuras orgánicas tridimensionales unidas por enlaces de hidrógeno (HOFs).

A lo largo de esta Tesis, se ha trazado una evolución progresiva desde redes porosas bidimensionales hacia materiales en 3D. En los Capítulos 2 y 3, se ha investigado el autoensamblaje de nuestras piezas moleculares sobre superficies. En el Capítulo 4, se ha avanzado hacia la preparación de películas multicapa con un espesor preciso. Finalmente, en el *Capítulo 5*, se ha enfocado el estudio sobre monómeros fusionados de forma covalente para la formación de sistemas de tipo HOF por crecimiento vertical, de forma ordenada y orientada hacia la creación de materiales cristalinos tridimensionales. A través de esta estrategia, son únicamente las interacciones fuertes y direccionales de tipo Watson-Crick las que guían el autoensamblaje en dos dimensiones (Figura 154). De esta forma, se puede alcanzar un control excelente de la porosidad del material, tanto de su naturaleza como de sus dimensiones.

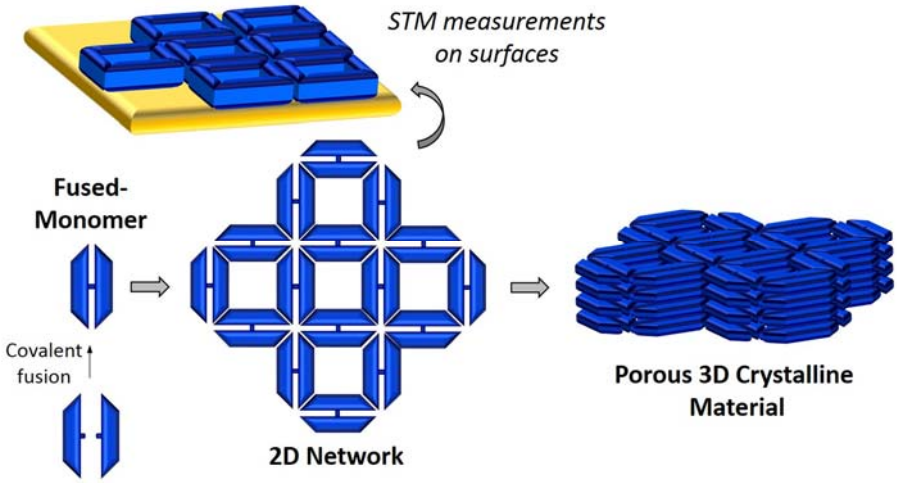


Figura 154. Representación esquemática de los objetivos abordados en el Capítulo 5.

Se ha dedicado mucho esfuerzo al diseño de estos materiales de tipo HOF mediante la construcción de diversos modelos moleculares teóricos. Se han considerado numerosos monómeros fusionados, tanto simétricos como no simétricos, con diferente grupo funcional de unión (reversible o irreversible, rígido o flexible, de difeent longitud…) y siempre basados en la estructura general de nuestros sistemas. Los dos diseños finalmente seleccionados son isoestructurales y cuentan con un primer conector robusto de tipo ftalimida, y con una segunda unión reversible basada en un éster borónico (Figura 155). Para preparar el primer derivado, se ha funcionalizado el bloque central con un grupo ftalimida deficiente en electrones, de forma que se pueda acceder al monómero fusionado **GC_GC1** tras reacción de transimidación con 1,4-diaminobenceno. Para el segundo diseño, se ha incorporado un grupo diol en la molécula para que, al añadir ácido borónico, esta

condense en **GC_GC2**. Ambos monómeros fusionados han sido dotados de cadenas alquílicas de una longitud determinada para que éstas ocupen parcialmente las cavidades no específicas que quedan entre tetrámeros cíclicos. Asimismo, las moléculas **GC_GC1**, **GC_GC2** y otros monómeros fusionados derivados de estas estructuras, servirán no sólo para dar lugar a diversos sistemas de tipo HOF, sino también para la producción de otras estructuras, como redes bidimensionales de tipo cadena o cinta. Estos sistemas permitirán acceder a redes bidimensionales robustas, dotadas de poros diseñados a medida, que pueden ser de forma cuadrada o rectangular, y que tras apilamiento, podrían producir materiales cristalinos porosos.

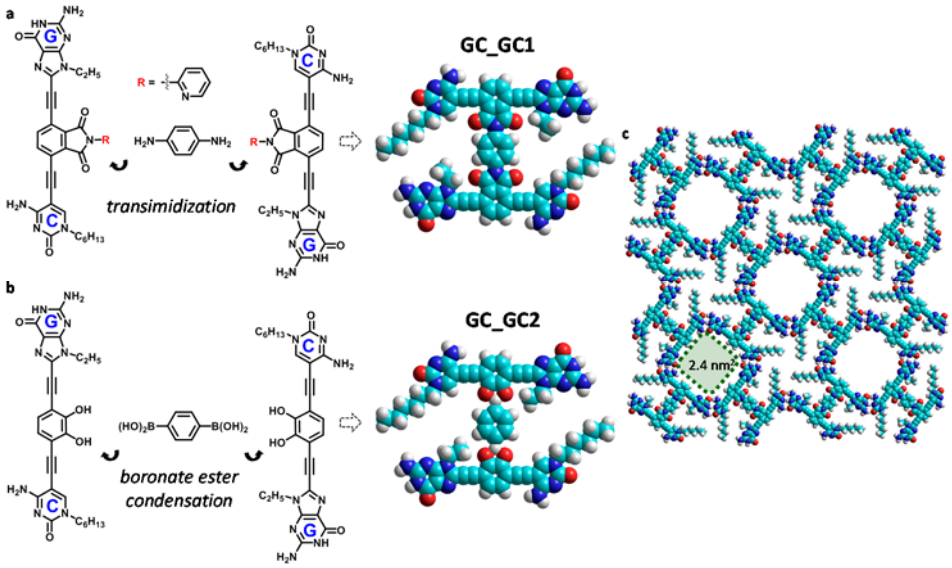


Figura 155. Estructuras de monómeros fusionados consideradas en el Capítulo 5 para la formación de sistemas de tipo HOF basados en uniones (**a**) ftalimida y (**b**) ester borónico. (**c**) Modelo de la capa bidimensional formada donde los poros no específicos han sido parcialmente bloqueados con cadenas alquílicas.

En definitiva, se ha desarrollado una estrategia novedosa, versátil y basada en el autoensamblaje para la nanoestructuración específica de superficies que sean capaces de reconocer moléculas en función de su tamaño, forma o estructura química. Una amplia colección de precursores ha sido sintetizada para que, a través de los acoplamientos de Sonogashira adecuados, se puedan preparar las moléculas deseadas basadas en las bases nitrogenadas del ADN. Estos monómeros son capaces de reconocerse entre sí para dar forma a tetrámeros cíclicos, que a su vez permiten acceder a interesantes redes nanoporosas hechas a medida. Con estas sencillas piezas moleculares, que pueden ser moldeadas a voluntad, una gran variedad de experimentos pueden ser diseñados para su estudio en diferentes condiciones. Con este sistema molecular tan fascinante, se pretende acceder a un nivel superior en el control de la nanoestructuración en 2D aplicada al reconocimiento molecular.